GENETIC AND STRUCTURAL IMPLICATIONS OF BACTERIOPHAGE P1 ASSEMBLY AND INFECTION

A Dissertation

by

MIGUEL F. GONZALES

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jason J. Gill
Committee Members,	Jennifer K. Herman
	Hays Rye
	Ryland F. Young
Head of Department,	Zachary Adelman

December 2022

Major Subject: Genetics

ABSTRACT

Coliphage P1 has been involved in major advances of molecular and microbiological tools. Nonetheless, modern molecular and microscopy applications have not been applied to further understand bacteriophage P1. Structural work to understand the P1 virion was conducted in the 1980's but has not been pursued since then. Furthermore, while the defense against restriction system of P1 has been observed to be packaged into the P1 virion in the stepwise manner, little is known about the genetic underpinnings of the anti-restriction system. In this work, mass-spectrometry was used to better understand P1 structural proteins. A combination of mass-spectrometry and bioinformatics suggested an appropriate role for P1 proteins identified by mass-spectrometry. I also report on the effects genetic manipulation have on the defense against restriction system, and to identify key motifs that are responsible for the anti-restriction activity. In addition, I report on the capsid localization signal of a key anti-restriction enzyme and demonstrate the capsid localization signal is sufficient to package mCherry into the P1 capsid. Finally, I report on a novel prong protein located on the P1 central spike. This prong protein is encoded by *upfB* and the lack of the prongs from the P1 virion result in an observable difference in infection of *V. cholerae* and particular *Enterobacteriaceae.* This work provides a foundation for the P1 structural identity as well as a basis to explore the defense against restriction system in further detail. Additionally, the identification of the P1 prong and its open reading frame can be used to identify homologous genes in other bacteriophage backgrounds.

DEDICATION

I dedicate this work to my grandparents Gloria Fernandez Noriega, Librardo Gonzales, and Consuelo Torrez Gonzales. They always believed in the value of higher education and pushed me to pursue just that.

ACKNOWLEDGEMENTS

I would like to first thank my parents Hilda N. Gonzales and Oscar Gonzales for without them I would not be where I am at. Their continuing encouragement and upbringing have provided me with a strong support and value system.

I am eternally grateful to my undergraduate mentors Daniele Provezano and Boris Ermolinsky. Their nurturing during my early adulthood was invaluable as it taught me how to think and to appreciate the scientific work that I have conducted. Without their guidance I am certain I would not have the peace of mind I have today.

I would like to thank Marcella Cervantez for being my first mentor in graduate school. She played a valuable role in my early graduate education, and I am sure no one else could do a better job.

I would like to thank Justin C. Leavitt for being the first friend I made in a laboratory setting in graduate school. Justin shared with me his immense knowledge on phage history and his obsession with phage structure and DNA packaging. I will always be grateful for the late night talks we had in lab while we were both working diligently on our projects.

I would like to thank the friends I've met while pursuing my work at Teas A&M. Their friendship has given me an outlet during the trying times of graduate school.

I would also love to thank my fiancé Ishita Chandel for if it were not for her, I would not have discovered the novel P1 prong. I am grateful for her love and support during my time earning my degree.

Finally, I would like to thank my committee chair, Dr. Jason J. Gill, and my committee members, Dr. Herman, Dr. Rye, and Dr. Young for their valuable guidance and support through the course of this research.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Jason J. Gill [advisor], Jennifer K. Herman of the Department of Biochemistry and Biophysics, Hays Rye of the Department of Biochemistry and Biophysics, and Ryland F. Young of the Department of Biochemistry and Biophysics.

The pBAD24g construct identified in chapter II, single knockouts identified in chapter II, the complementation vectors identified in chapter II, and the mCherry fusion construct identified in chapter III were made by Dr. Denish Piya. Edman sequencing presented in chapter II was conducted by UC Davis. Mass-spectrometric analysis was conducted by the Texas A&M Department of Chemistry. Cryo-EM grid preparation, screening, and data acquisition was completed by the Pacific Northwest Center for Cryo-Electron Microscopy. Alpha Fold prediction was conducted by Professor Petr Leiman.

All other work conducted for this dissertation was completed by the student independently.

Funding Sources

Graduate study was supported in part by the National Science Foundation Graduate Research Fellowship Program (DGE:1746932).

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Of life on earth, one of the simplest forms is the bacteriophage (phage). Phages are obligate parasites of bacteria. The phage population on earth is estimated to be 10^{31} at any point in time [1]. Phages are composed of a proteinaceous shell which houses nucleic acid capable of being transmitted to a susceptible host. Nearly 96% of described phages are identified as tailed phages, which are characterized by double stranded DNA packaged into icosahedral capsids attached to a tail that serves as a tool for infection [2,3]. Tailed phages have traditionally been further subdivided based on their tail morphology: myoviruses which have long, rigid, contractile tails, siphoviruses which have long, flexible, non-contractile tails, and podoviruses which have a stubby non-contractile tail (Figure 1). The hallmark phages for each morphology are phages T4, lambda, and T7, respectively as foundational work in those morphologies was first addressed in these phages.

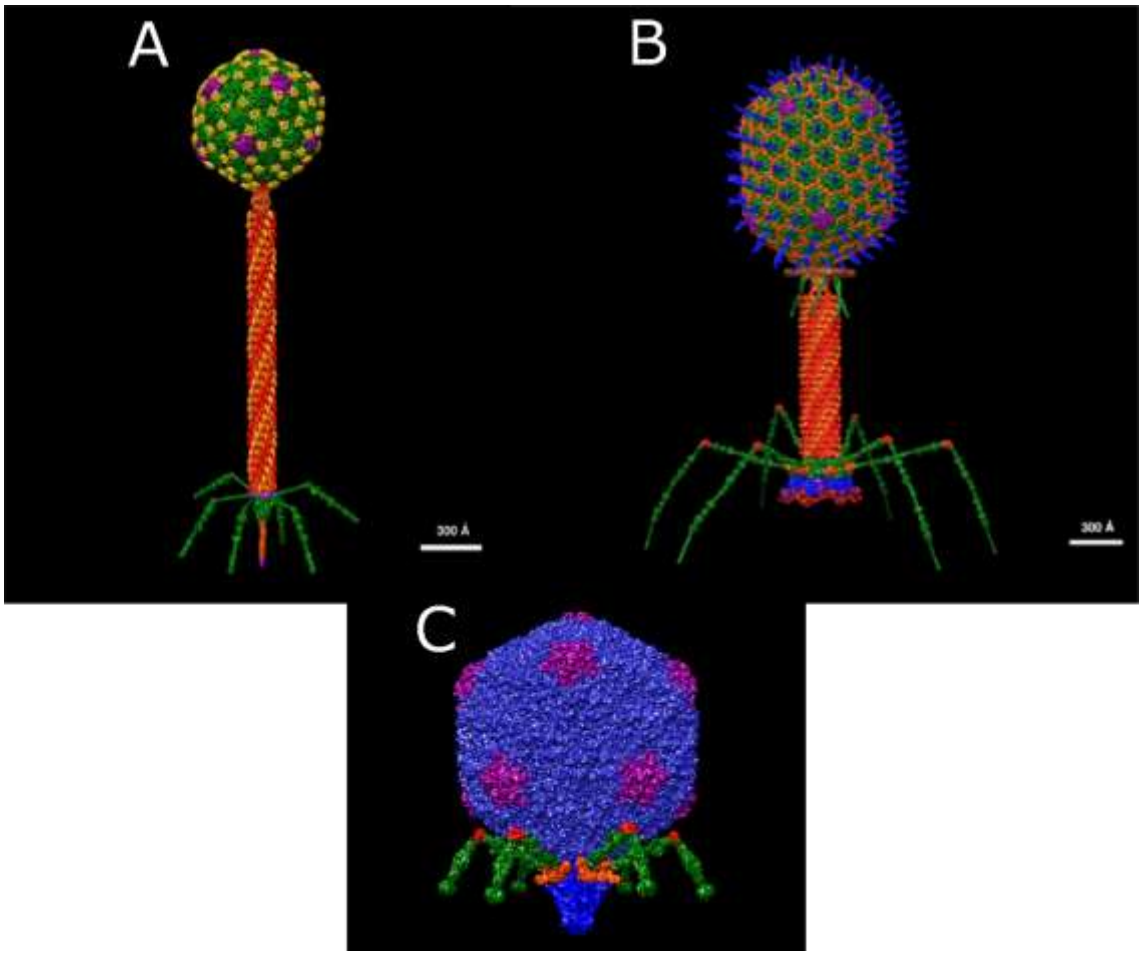

Figure 1 Morphology of Siphoviviruses, Myoviruses, and Podoviruses. A) A USFC Chimera reconstruction of phage lambda at atomic resolution. B) A USFC Chimera reconstruction of phage T4 at atomic resolution. C) A USFC Chimera reconstruction of phage T7 at atomic resolution. Image recievd from Wikimedia commons.

While phage morphology is a useful way to characterize phages into three broad groups, phage genomes are more complex than this phenotype. Indeed, the ancestry of phages is difficult to ascertain due to the rich recombinatory environment that phages produce when replicating leading to 'genetic mosaicism' – a term which describes "modular" sequence similarity observable when two related phage genomes are compared [4]. All phages adsorb to a susceptible bacterial host and transmit their nucleic

acid to the host bacteria [5]. With a frame of reference of phage of the three morphologies mentioned above, the infection cycle starts with the detection of a bacterial host by the phage's tail fibers. Once the phage has absorbed to the host cell, the phage is able to transmit its nucleic acid into the host cytoplasm.

Phages can exhibit two defined lifestyles termed virulent and temperate [6]. Virulent phages immediately begin to produce viral progeny and undergo the lytic pathway while temperate phages do not immediately begin to produce viral progeny upon infection. However, if the accumulation of a master repressor reaches above a threshold, then phages incorporate their viral genome into the host chromosome. When incorporated, the phage genome is replicated during host DNA replication and inherited by daughter cells [7]. In this way, the temperate phage passes through multiple generations until the master regulator experiences a stressor that triggers the prophage to resume the lytic pathway. The lytic pathway is characterized by viral progeny production and concluded by lysis. While the specifics of lysis will be discussed in another section, the result of lysis is the disruption of the bacterial host cell envelope which enables the release of viral progeny from the host cytoplasm. Lysis occurs in virtually all double stranded DNA phages even though the morphology and genome architecture vary. The completion of lysis of the bacterial host liberates the viral progeny from the bacterial cytoplasm. In this way, the viral progeny can infect a new host and repeat the cycle [6].

Phage Lifestyles

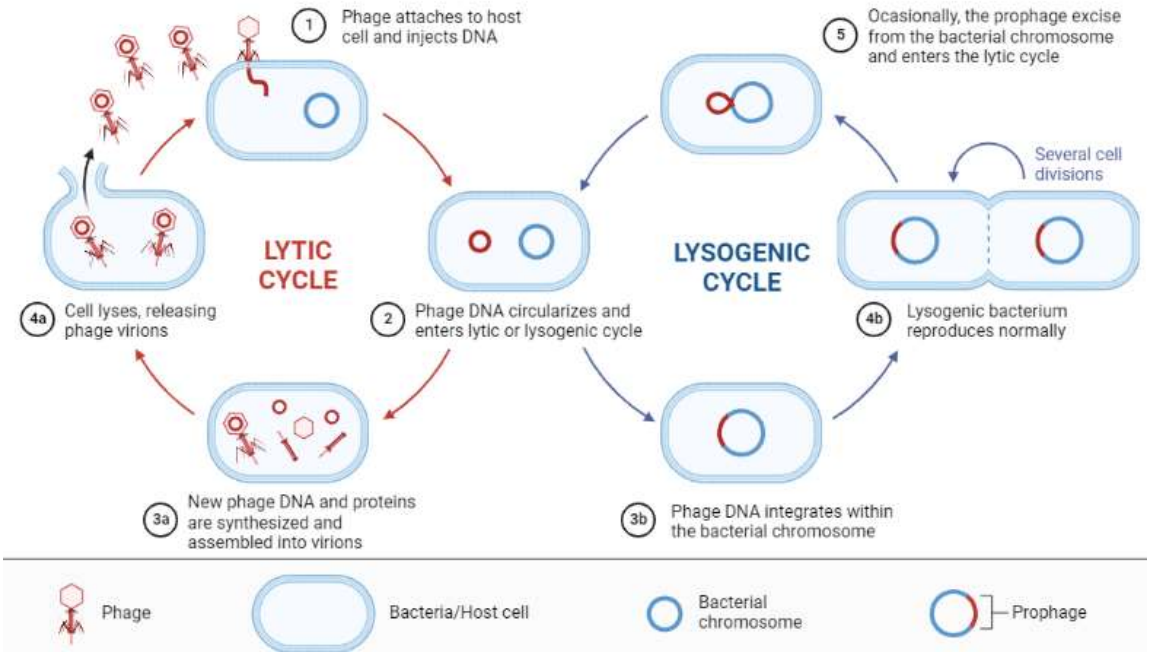

Figure 2 Diagram of the life cycle of a *Caudovirales.* The stages of the life cycle are detailed in this section. The graph combines multiple steps into one stage. The lytic and lysogenic cycle diverge at step 3. 3a describes a step in the lytic cycle, while 3b is a step in the lysogenic cycle. The red circle observed in step 2 is the phage circularized phage chromosome. Adapted from "Lytic and Lysogenic Cycle", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

Bacteriophages have three life cycles – virulent, temperate, and chronic infection [8]. Chronic infection is exclusive to filamentous phage and encompasses the filamentous phage life cycle wherein viral progeny are secreted from the host cell without lysing the cell membrane [9]. There are five stages in the phage lytic cycle: adsorption, nucleic acid injection, assembly of progeny virions, progeny virion release from the host cytoplasm, and subsequent infection (Figure 2) [10]. Again, adsorption describes the physical interaction of one object adhering to the surface of another object. In *Caudovirales*

phage infection, the adsorption step of the life cycle describes the step in which the phage attaches to the host outer membrane. This step requires the specific recognition of host surface receptors such as the lipopolysaccharides and other molecules such as flagella [11] . The recognition of host receptors by the phage causes an irreversible event which allows the phage to penetrate the bacterial membrane and eject the nucleic acid into the host cytoplasm. The cell envelope is degraded by phage-encoded lysozymes which hydrolyze 1,4-β-linkages between N-acetylmuramic and N-Acetyl-D-glucosamine polysaccharide within peptidoglycan present in the periplasmic space [12]. Once the viral nucleic acid is in the host cytoplasm, the phage will either undergo the lytic or the temperate cycle. If the expression of the master repressor is high enough to suppress the lytic cycle, then the phage establishes lysogeny and is replicated during synthesis phase and inherited by daughter cells. In response to an intracellular signal, the induction of the lytic cycle occurs, initiated by the dissociation of the master repressor from prophage operators. In the late stage of the lytic cycle, the progeny virions exit the host cytoplasm with the aid of lysis proteins known as endolysins, holins, spanins, and, in some cases sar-endolysins [13]. The role of the endolysins is to degrade the peptidoglycan while the holin and spanins disrupt the destabilized bacterial membranes, respectively. Together, endolysins, holins, and spanins enable the progeny virions to exit the host cytoplasm and infect other cells [13].

Transduction and Lysogenic Conversion

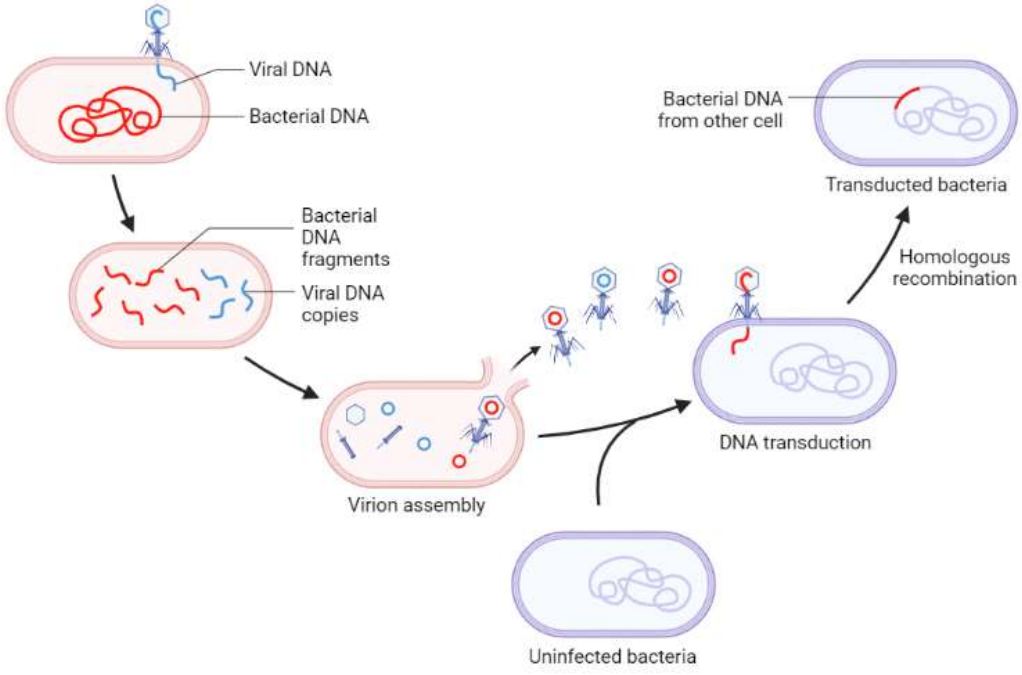

Figure 3 Diagram depicting transduction via a myophage. The diagram stars at the top left interaction depicting an adsorbed phage injecting its DNA into a bacterium. Next, bacterial and phage DNA are both confined in the bacterial cytoplasm. Then, the bacterial DNA is able to be packaged into progeny virions as identified by phages packaged with red circles. This phage is capable of transduction. Next, the phage packaged with bacterial DNA infects a new bacterium. Adapted from "Bacterial DNA Transduction", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

Phage mediated horizontal exchange is a major driver for biological diversity [8]. A major phage-mediated driver for genetic transfer among bacteria is transduction [8]. Transduction occurs when a progeny virion packages bacterial DNA (Figure 3). The subsequent infection of this phage introduces the packaged bacterial DNA into a new cellular environment. Transduction can be generalized or specific. In generalized transduction, random fragments of the bacterial DNA are packaged into progeny phage.

The random bacterial DNA is introduced into a new bacterial cytoplasm and is either integrated into the chromosome by recombination, exist as an extra chromosomal element, or degraded [8]. Specialized transduction describes the packaging of DNA that is adjacent to the prophage integration site in the bacterial chromosome [14].

When a prophage integrates itself into a host and confers the bacterial host with a beneficial phenotypic change, the result is called lysogenic conversion [15]. A lysogenic conversion results in a symbiotic relationship between the prophage and the bacterial host wherein the bacterial host receives a fitness advantage by increased pathogenicity and the prophage is able to replicate and propagate its genome along to another generation [16]. In the case of CTXΦ in *V. cholerae*, two proteins, toxin co-regulated pilus and cholerae toxin increase *V. cholerae* pathogenicity. The symbiotic relationship between phages and bacteria can enable the bacterial host to survive in new niches [17]. Examples are the ability to form biofilms and aid biofilm dispersal, proteins that encode toxins to combat competing organisms, and adhesions which allows bacteria to invade a new niche [18]. Additionally, a common phenotype that is imparted by lysogenic conversion is the resistance to clinically relevant antibiotics such as β-lactamses such as the gene bla_{TEM}, $bla_{PSE\text{-}1}$, $bla_{OXA\text{-}2}$, and bla_{PSE} found in *E. coli*, *C. jejuni*, *P. aeruginosa*, and *S. enterica*, respectively [19-21]. Moreover, if the prophage does not encode for virulence factors, the prophage may affect the production of bacterial toxins and regulate gene expression – a phenomenon known as active lysogeny [22]. Further, prophages can shape microbial environments through lysis events [23]. Prophages can transmit virulence factors to new bacterial hosts allowing their hosts to invade new niches,

survive in novel environments, alter gene expression, and even disrupt gene expression. Each of the examples can have profound effects on ecological structure [23].

Phage Capsid Structure and Assembly

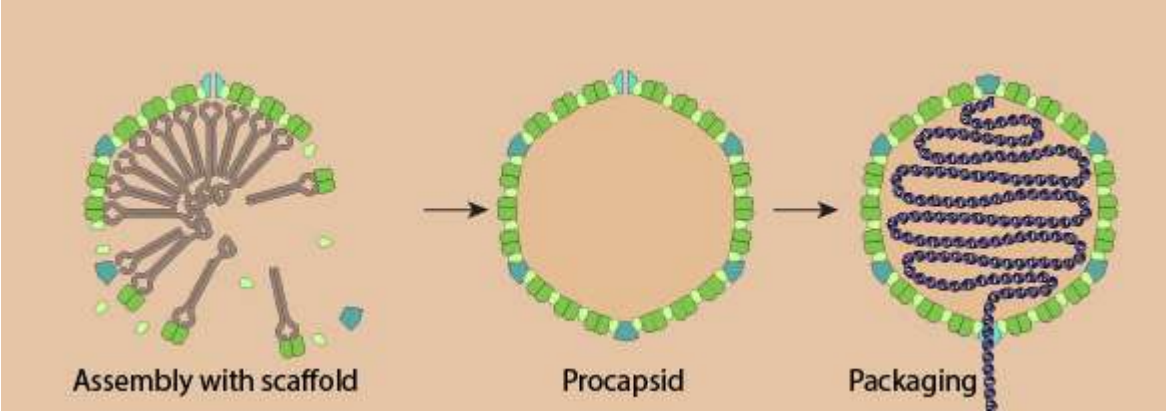

Figure 4 The simple depiction of capsid assembly. The illustration on the left is of procapsid assembly by joining of the major capsid protein and the pentameric proteins via a scaffold protein. Next, the image in the middle displays the scaffold proteins dissociated from the assembled procapsid. Finally, the figure on the right shows the DNA pumped into the capsid by the terminase complex that is bound to the portal vertex. Image retrieved from https://viralzone.expasy.org/140?outline=all_by_species

It has been observed that a specific β-barrel fold was a common feature in icosahedral viral capsids [24]. This 'jelly roll' β-barrel was identified in different viruses which infect nearly every kingdom of life [25]. The observation of the common fold between eukaryotic and prokaryotic viruses suggested a common evolutionary origin [26]. Additionally, it was found that *Herpesviridae* (herpesviruses) and *Caudovirales* share a characteristic fold that is unique to these two lineages which suggests a common ancestry [27]. Taken together, the study of phages provides insight for not only phage biology but may also be used to understand clinically relevant viruses as well. The assembly of virions of the order *Caudovirales* occurs in a similar pathway [28]. Put

simply, assembly of myophage and siphophage follow as similar pathway: the capsid and tail are independently made and then joined to form a virion. Since each pathway is independent of the other, each assembly pathway can be analyzed independently.

A survey of 5,500 phage micrographs found capsids of the order *Caudovirales* in a variety of sizes ranging from 400 to 1700 angstroms in diameter [29]. Additionally, most *Caudovirales* observed by electron microscopy have icosahedral capsids while nearly 15% have "prolate" heads. Prolate heads are icosahedrons elongated along the 5-fold axis coincident with the axis of the phage tail [29].

The capsid contains linear double-stranded DNA which is packed to an internal pressure of nearly 20 atmospheres [30]. Capsid assembly nucleates at a unique pentameric vertex comprised of the dodecameric portal protein. The portal protein forms a channel for packaging the phage genome during capsid assembly, and allows the genome to exit during infection [28]. The portal vertex polymerizes with a scaffold protein and the major capsid protein to form a procapsid (or prohead) (Figure 4) [28]. Thus, a procapsid is composed of the portal protein, and the internal scaffolding core, surrounded by the major capsid protein shell [31]. At this point, when compared to the mature capsid, the procapsid lacks an icosahedral shape, is smaller, and has a rounded shape. It is important to note that not all phages utilize a separate scaffolding protein. For example, phage HK97 does not have a dedicated scaffold protein, but instead the N-terminal region of the major capsid protein acts as a scaffolding [32]. With the exception of phage P22, the procapsids also contain head maturation proteases which serve to degrade the internal scaffolding core either partially or completely to allow for the

scaffolding core to vacate the procapsid. This frees the internal cavity of the procapsid for the genomic DNA [33]. The procapsid shell is then packaged with DNA through the portal vertex by the ATP-dependent activity of the terminase motor complex [34]. While the procapsid is being filled with genomic DNA, the procapsid undergoes a structural rearrangement to transform into a mature capsid. After the mature capsid is formed, many capsids are observed to have decoration and/or stabilization proteins attached to their outer surface during the final stage of maturation. Once the procapsid has been filled with genomic DNA by the terminase motor complex, the complex dissociates from the portal vertex. Finally, head completion proteins are attached to the capsid which serve to seal the portal vertex and provide a site for the tail to nucleate or bind to the capsid as is the case for podophages, and sipho-and myophages, respectively.

Internal Capsid Proteins

Notable *Caudovirales* phages are understood to package proteins inside their capsid. Proteins packaged inside the phage capsid range in function. In the case of myophage T4, Alt is packaged into the capsid and modifies the host's RNA polymerase resulting in increased binding to phage early promoters. Additionally, the protein gp2 is also packaged inside the T4 capsid which protects the phage DNA from being degraded by host exonuclease restriction enzymes [35]. Further, T4 contains more than 1000 internally packaged copies of 3 unique small proteins [35]. These proteins are localized to the T4 prohead by a capsid-targeting sequence [36]. Specifically, internal protein I also protects the phage genome from degradation by bacterial endonucleases through direct protein-protein interaction [37]. The internal proteins are incorporated into the T4

prohead via a capsid localization signal that targets proteins to the procapsid core [35]. These small proteins are not essential for capsid assembly or for the stability of the packaged DNA [38]. Interestingly, it has been suggested that the dimensions of internal protein I are such that it can be ejected from the capsid. Thus, enabling immediate defense against restriction [37]. Curiously, in the case of T4, the internal protein III-β-galactosidase fusion exhibits activity within the host bacterial cell [39]. It is apparently impossible for β-galactosidase to be packaged inside the T4 capsid in a folded and functional state as the 540-kDa tetramer-dependent activity requires host chaperones to fold. This is evident by increased activity of β-galactosidase when bacteria are used to harvest the protein than activity observed from protein extracted from purified phage heads. Taken together, it is reasonable to suspect that the unfolded β-galactosidase is ejected, then refolding and multimerization takes place within the bacterial host cytoplasm [39]. In podophage T7, an internal core is assembled on the portal protein [40]. The internal core of T7 is composed of multiple copies of gp14, gp 15 and gp16 which exit the capsid during infection to form an extension of the phage tail which traverses the host's membranes and periplasmic space [41]. While packaged, gp14, gp15, and gp16 are arranged in three layers with 12-,8- and 4-fold symmetry, respectively [40]. In podophage N4, nearly four copies of gp50, one of N4's RNA polymerase, is packaged in the phage capsid [42]. Like the internal core of T7, gp50 is located at the base of the N4 inner core, above the portal, such that the RNA polymerase is positioned close to the internal entrance of the portal channel [42]. N4's gp50 is injected into the bacterial host during infection. Gp50 is understood to transcribe early phage promoters, and simultaneously

aid DNA ejection by physically pulling the remaining genome from the capsid into the host cytoplasm [43]. The binding of N4's RNA polymerase to an N4 protomer is also understood to be aided by bacterial host proteins, specifically, DNA gyrase and single-stranded DNA-binding protein [44] [43]. Additionally, N4's RNA polymerase, much like DarB, is also extremely large (~380 kDa) with the central domain responsible for RNA polymerase activity, an N-terminal domain that the required for ejecting the first 500 bp of the N4 genome into the host cytoplasm, and a C-terminal domain that is required for the localization of gp50 to the N4 capsid [45]. Much like T7 internal proteins, cryo-EM reconstruction shows N4's gp50 to be localized near the portal protein and forms a foundation for the internal core [42]. Another podophage with internally packaged proteins is phage P22. P22 harbors gp7, gp16, and gp20 in the mature capsid [46]. These P22 proteins are also ejected into the host cytoplasm upon infection [46]. Finally, in the giant bacteriophage φKZ, a large internal protein with dimensions of nearly 1050 Ǻ long and 240 Ǻ wide is packaged [47]. The two ends of the internally packaged protein are anchored to hexameric capsomers to provide stability. Additionally, the 100-200 copies of 5 major proteins are packaged inside the capsid of φKZ [35].

Phage Tail Assembly

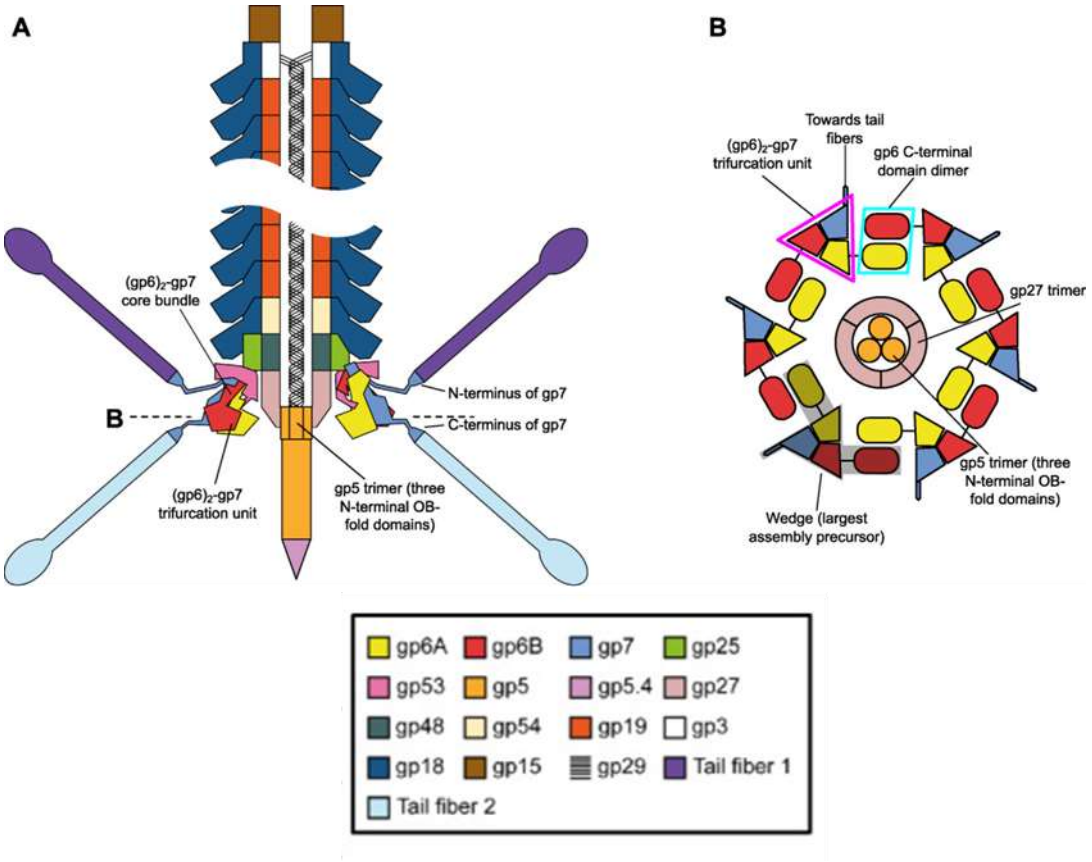

Figure 5 A schematic of the assembled T4 tail. The tail is color coded such that each protein is identified as a different color. Shapes of the same color are copies of the same protein. The colors within A and B correspond to a color within the legend. A) The tail fibers are shown in two colors to simplify the multicomponent tail fiber network (Tail fiber 1 and tail fiber 2). Each are attached to the N- and C-terminus of gp7. The dashed horizontal line identifies the position of the cross-section shown in panel B. B) Gp6-and gp7 trifurcation unit and the gp6 C0terminal domain dimer are identified by magenta and cyan lines respectively. The largest assembly precursor wedge complex is highlighted with a semitransparent grey shape. The figure is figure 1 from "Contractile injection systems of bacteriophages and related systems" https://onlinelibrary.wiley.com/doi/10.1111/mmi.13921

As stated previously, sipho- and myophage tails are long structures. In phage T4, the baseplate is comprised of about 140 polypeptide chains of at least 16 proteins [48]. Broadly, the myophage baseplate is composed of a central hub and 6 wedges. The T4

hub is composed of gp27, gp5 and gp5.4 while the wedges are composed of gp6 and gp7. While most of the T4 baseplate exhibits 6-fold symmetry, the central hub of the baseplate is composed of the trimeric gp27 and gp5 proteins. The hub nucleates from a trimer of the tape measure protein. The tape measure protein extends from the baseplate and establishes the length of the T4 tail. Studies in T4 demonstrate that truncations and gene duplication of the tape measure protein alter the length of the phage tail in negative and positive fashions, respectively [49]. Thus, the length of a phage tail is proportional to the size of the tape measure at 1.5 angstroms length per residue. The tape measure protein servers as a contact for the tail tube to polymerize starting from the baseplate. Once the tail tube reaches the end of the tape measure protein, a hexameric tail tube terminator protein associates with the top of the tube [50]. In the case of myophages, the contractile sheath is assembled around the tail tube. The polymerization of the tail sheath nucleates at the baseplate and polymerizes until it reaches the end of the tail tube. The tail sheath subunits polymerize into a hexameric stack of rings creating a 6-fold symmetry [48]. The sheath is terminated by the hexameric sheath terminator protein which binds to the top of the tail tube terminator and contacts the last segment of the tail sheath (Figure 5 gp18) [51].

Myophage T4 gp13 and gp14 attach to the portal vertex of the capsid and then bind directly to gp15 of the tail which caps the tail sheath and tail tube [52]. It is presumed that the gp13-gp14 complex has sixfold symmetry and binds to the gp20 dodecamer [53]. After the tail and head are bound, Wac proteins form the whiskers of phage T4. To complete the virion assembly, six long tail fibers are attached to the

baseplate. Tail fiber attachment occurs in multiple steps starting with the hinge region of the tail fiber first binding to the C-terminal end of the whisker, thus orienting the tail fibers so that the N-terminal end of gp34 binds to gp9 in the upper edge of the baseplate. Additionally, gp63 facilitates tail fiber attachment [54]. Once the tail fibers are attached, an infectious virion is assembled (Figure 1B).

Phage Adsorption and DNA Ejection

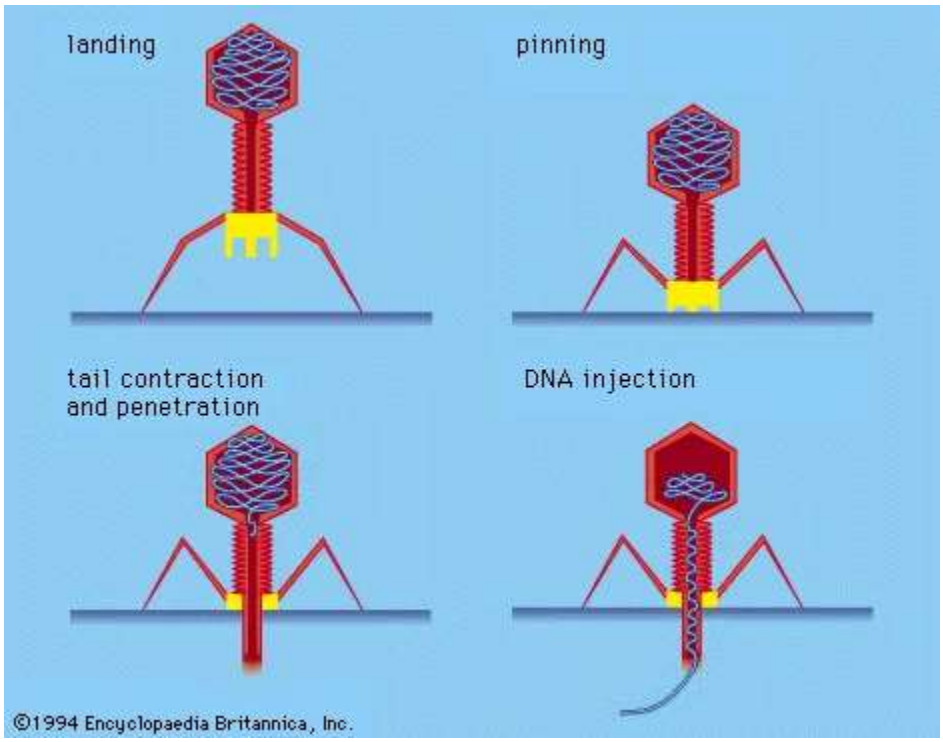

Figure 6 Steps illustrating the adsorption of a phage onto the host membrane and the ejection of DNA into the host. "Landing" depicts a myophage adsorbed to a host's outer membrane. "Pinning" depicts the contact of the central spike to the outer membrane. Pinning is followed by the contraction of the tail which penetrates the hosts cell envelope. Finally, the DNA packaged within the phage capsid is ejected into the host cytoplasm. Image retrieved from https://www.britannica.com/science/bacteriophage

The initial interaction between a bacteriophage and its bacterial host is the specific binding between the phage and the bacterial surface receptors – a process known as adsorption (Figure 6). Phage adsorption occurs in two phases wherein a reversible step involves the binding of the phage to the primary receptor of the bacteria. This is followed by an irreversible step where the secondary receptor is recognized by the phage [5]. Phage components involved in adsorption are class specific. The components include the tail fibers, tail spikes, tail needle, baseplate proteins, and bacteriophage encoded

lyases [55]. The configuration of the tail fibers of bacteriophage T4 must be pointing away from the body of the virus [56]. This orientation of the T4 phage is considered "active". Studies have identified L-tryptophan to be an important cofactor for T4 to achieve an active state [56]. Endogenous tryptophan residues located on the baseplate or tail sheath interact with L-tryptophan leading to an active tail fiber configuration [56]. Once three long tail fibers are bound to the host cell receptors, the fibers signal the baseplate through gp9 that successful binding has occurred [57]. The baseplate is brought to proximity of the cell surface and the baseplate initiates contraction of the tail sheath causing the rigid tail tube to pierce the outer cell membrane using the pointed needle formed by gp5 [57]. Gp5 dissociates from the T4 virion and digests the peptidoglycan layer to create an opening to which the tail tube can contact the cytoplasmic membrane of the bacterium [57]. The contact of the tail tube with the cytoplasmic membrane triggers the release of the phage DNA into the host through the tail tube [57]. The internal protein I of phage T4 is injected to protect the DNA from *E. coli* gmrS + gmrD glucose modified restriction endonuclease [58]. Protein I has a novel protein fold consisting of two beta sheets flanked by N- and C-terminal alpha helices which bind gmrS/gmrD proteins and inhibit endonuclease activity [58].

Bacterial Host Defenses

Bacteria are able to evade phage infection in different ways. In general, there are three classes of resistance mechanisms: receptor adaptations, host defense systems, and phage-derived phage defense systems [59]. Receptor adaptations result in decreased phage adsorption [59]. There are 5 subclasses within host adaptations that result in phage

resistance: point mutations, receptor masking proteins, outer-membrane vesicles, increased productivity of extracellular matrix, and phase variation. Point mutations may results in a loss or change in phage receptors as demonstrated by *E. coli* challenged with phage U136B wherein *E. coli* gene *tolC* accumulated mutations as a means of resisting phage infection [60]. *Acinetobcter baumannii* mutates genes *gtr29* and *gpi* which are involved in the biosynthesis pathway of capsular polysaccharides to elude infection by phages øFG02 and øCO01, respectively [61]. Additionally, *Listeria monocytogenes* losses cell wall teichoic acid rhammosylation resulting in phage resistance [62]. Further, deletion of genes encoding for bacterial receptors experienced a decrease in phage infection [63].

Bacteria also encode proteins that mask phage receptors on the cell surface. A prime example of this is the F plasmid-encoded protein TraT which binds to the surface exposed regions of the outer membrane porin OmpA in *E. coli* [64]. While phage infection is not completely inhibited by TraT, it is assumed that TraT only binds to a segment of OmpA therefore decreasing the exposure of the recognized protein required for T-even phage binding [64]. Bacteria are also able to create "decoys" which phage adsorb to. *E. coli yieM* mutants are a hyper-vesiculating mutant of *E. coli* while still maintaining membrane integrity [65]. Vesicles from *E. coli yieM* mutants were shown to decrease bacteriophage T4 infection and electron microscopy shows that T4 was bound to outer membrane vesicles [66]. Further, diffusion experiments with *Pseudomonas aeruginosa* suggested that bacterial strains that disable phage diffusion were more resistant to phage infection [67].

Bacteria have also evolved defense islands which encode defenses against mobile genetic elements, which includes phage [68]. A wide-spread example is that of the phage defense system called restriction-modification. In a restriction-modification system, a methyltransferase methylates the bacterial genome at specific sites protecting it from cleavage by the restriction endonuclease which recognizes unmodified DNA and cleaves within or at a distance from the recognition site [69]. The four classes of restriction-modification systems are based on the characteristics of their specific components [69]. Type I restriction-modification systems consist of a protein complex with three subunits, each with a distinct activity – the methyltransferase, restriction endonuclease, and the specificity subunit. The specificity subunit recognizes a target sequence within DNA. The target sequence also serves as the cleavage site of the restriction endonuclease. Type II is the most abundant and is commonly used as a tool in molecular biology. Type II restriction modification systems are composed of two independent methyltransferase and restriction endonuclease proteins. The restriction endonuclease of type II restriction modification systems cleaves at the site of recognition which makes them an invaluable tool for molecular cloning. Type III restriction modification systems are similar to type II in that there are two independent methyltransferase and restriction endonuclease proteins However, in type III each subunit can function independently. Type IV restriction modification systems cleave modified DNA differing from type I and II systems which target unmodified DNA. T bacterial genomes that encode type IV R-M systems are unmodified [70].

Another host phage defense system is abortive infection which is described as a series of mechanisms that lead to bacterial cell death [71]. Abortive infection is thought to serve as a strategy used by bacteria to protect the population by programmed suicide of the infected bacterium [71]. This system has been explored in bacteria such as *E. coli* and *V. cholerae* for example [72] [73]. Abortive infection systems commonly rely on a toxin-antitoxin mechanism wherein the balance between a stable toxin and an unstable antitoxin determine the fate of the cell [74]. The antitoxin is continually degraded but replenished so bacteria growth isn't affected by the toxin. However, when a phage infects a bacterium, the bacterial cell can longer maintain levels of the antitoxin resulting in a disproportionate amount of toxin causing the death of the bacterium [74]. Toxin-antitoxin systems are understood to fall into the following two categories: protein-protein systems, such as the P1 Phd-Doc proteins [75], and RNA-RNA systems such as *hok* and *sok* [76].

Phage Overcoming Host Defenses

Phages have multiple mechanisms to combat bacterial defenses [59]. Phage overcome a bacteria's variation in cell surface receptors via mutations in receptor binding proteins [59]. These regions experience targeted mutation due to an error-prone reverse transcriptase [77].

To overcome the extensive extracellular network of lipopolysaccharides produced by bacteria, phages degrade these structures using depolymerases. These enzymes are located in the tail spike as is the case for phage T4 [54], or released in a soluble form following adsorption as is the case for P22 [78].

Phages have also developed mechanisms of escaping bacterial restriction-modification systems. One example of phage defense against these systems is the mutation of restriction sites from the phage genome resulting in the avoidance of host restriction enzymes [79]. Another example is how phage T4 employs glucosyl-hydroxymethylcytosine as a replacement for all cytosine bases [80]. Furthermore, phage T7 expresses the protein Ocr which mimics the structure of double stranded DNA sequestering restriction endonucleases. Phages can also avoid abortive infections by expressing their own antitoxin proteins as is the case for bacteriophage T4 [81].

Coliphage P1

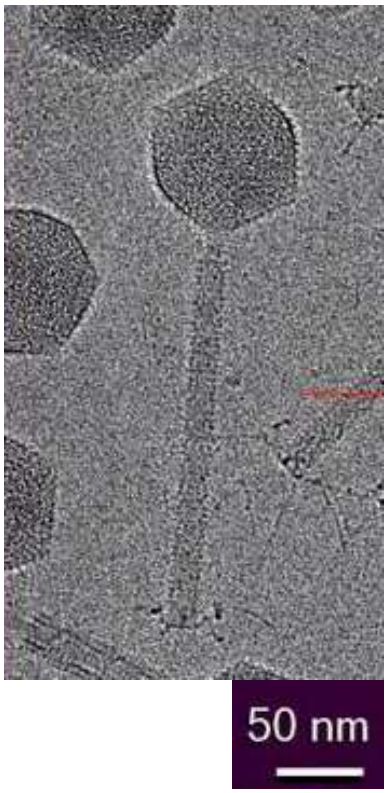

Figure 7 Cryo-electron micrograph of phage P1. Magnification is 75,000x in the presented micrograph. The micrograph is a cropped image from a micrograph used to screen P1 grids for particle density to confirm the value of the grid for data acquisition. The tail of phage P1 is about 120 nm while the head size is approximately 90 nm. The tail fibers are attached to the baseplate of P1 and can be readily observed in this micrograph.

The *E. coli* phage P1 (Figure 7) was identified in 1951 by G. Bertani while studying the lysogenic strain Lisbonne of *E. coli* [82]. The "P" used in "P1" is significant as Bertani identified three distinct plaque morphologies and named them P1, P2, and P3 [82]. At the time of study, phage P1 lysis was understood to be temperature sensitive [83]. In fact, it was observed that P1 would remain lysogenic when incubated at 20°C while induction of the P1 prophage took place at 37°C [83]. Early work on P1 described

P1's ability to transduce host DNA [84]. Not only was P1 able to transduce genetic characteristics from *E. coli* and *Shigella dysenteriae*, but P1 was able to transduce characteristics of λ prophage [84]. The frequency of transduction by phage P1 for *E. coli* strain K12 was reported to be approximately 10^{-6}. It is the transducing capability of P1 that enabled the mapping of the *E. coli* genome, enabling mapping of the *E. coli* genome [85].

P1 is unusual in that it does not integrate its genome into the host chromosome, but rather exists as an extrachromosomal element [86]. P1 was also observed to be unusual because it produced capsids of two different sizes [87]. One capsid is about 90 nm in diameter, while the other is about 60 nm in diameter. The two different head sizes were separable by cesium chloride density gradients [87]. It was further demonstrated that both particles contained DNA [88]. Though the small heads are not able to form plaques, they are able to kill the bacterial host and are capable of adsorbing and injecting DNA [87]. Additionally, there is a 'minute' capsid morphology capable of transferring DNA. The tail associated with the three capsid sizes are indistinguishable [87]. While capsid morphogenesis in understood in the context of essential genes in the case of phage T4 [89], phage P1 provides a unique model to explore the structural implication of proteins that are involved in defining the size of the phage capsid.

P1 was made chloramphenicol resistant during an investigation of transducing R factors with phage P1 [90]. This derivative of phage P1 was able to form plaques and confer anti-biotic resistance when establishing lysogeny. No detectable differences were observed between normal P1 and antibiotic resistance P1 in terms of density gradient

analysis in cesium chloride, the stability of lysogenization, ability to transduce chromosomal markers, and UV induced induction [90]. Amber mutation analysis was conducted on phage P1 [91]. The amber library was leveraged to elucidate regions of the phage genome important for lysis [92] and particle morphogenesis [93]. The amber library is now considered to be unusable (private communication with Robert Duda in 2019).

Phage P1 harbors an anti-restriction system, identified as a "defense against restriction" or Dar system [26]. Initially, it was observed that DarA and DarB both contributed to the defense against restriction system with DarA enabling protection from restriction enzyme EcoA and StySA, and DarB affording protection from EcoB, EcoK, StySB, StySP, and StySQ [94]. Protection of DNA was not specific to P1 DNA as any DNA packaged into the P1 capsid was protected [87]. It was also observed that when DarA is not available, capsid assembly is enriched in small heads [93]. Additionally, homologs of P1 protein DarB have been identified in many mobile genetic elements and has been implicated in facilitating horizontal gene transfer [95]. The study of the anti-restriction activity of P1 provides insight into basic evolutionary forces acting on prokaryotes furthering our understanding on horizontal gene transfer.

The P1 genome is described as T4 like as it is organized into gene modules related to their function [96]. Protein coding genes occupy 92% of the genome organized in 45 operons [96]. The master repressor for phage P1 is the C1 repressor which controls seventeen operons transcribed from σ^{70} promoters [96]. The *E. coli* RNA polymerase holoenzyme in combination with Lpa protein encoded by P1 control the late operons.

Since P1 is a temperate phage, it requires a set of genes to establish lysogeny and maintain itself as a prophage. Upwards of 20 genes are within operons that are not controlled by C1 and Lpa and are either predicted or known to be involved with lysogeny [96]. Maintenance of the P1 prophage requires the constitutive expression of the following factors; the C1 protein and C4 anti sense RNA which inhibits translation of the P1 anti-repressor Ant, and Lxc which is an auxiliary repressor protein [97]. The auxiliary Lxc protein enhances binding of the C1 repressor to its operator [97]. The translational repression exhibited by the C4 RNA blocks *icd* translation and in turn blocks transcription of *ant* by a *rho*-dependent terminator [98]. As stated previously, the master repressor of P1 is the C1 protein encoded by *c*1 within *ImmC* [96]. It has become common for the allele of *c*1, *c*1-100, to be used as it is thermosensitive [99]. It is suspected that the amino acid substitution G193C in C1 confers the thermosensitivity characteristic to the C1-100 protein [96]. C1 exists as a monomer in solution and binds to a 17 bp operator consensus sequence in an asymmetric manner which imparts directionality [100]. Another important protein for lysogeny appears to be Lxc. Lxc has been observed to strongly inhibit the ability of an anti-repressor, Coi, to dissociate operator-C1 complexes, so Lxc may help in lysogeny establishment [101]. Curiously, Lxc has also been suggested to increase the affinity of C1 to its operator and down-regulate the synthesis of C1 [102]. Taken together, Lxc may enable operators with weak affinity to C1 to prevail in low C1 environments, while also enabling C1 binding to promoters that exhibit a strong affinity for C1 [96]. Even though P1 lysogenizes at a low copy number [103], maintenance of P1 as an extrachromosomal element is only lost at a frequency of

nearly 10^{-5} per cell per generation [99]. The stringent frequency of inheritance is attributed to mechanisms of plasmid replication, plasmid partitioning, a site-specific recombination system, and P1's plasmid addiction system.

DNA replication of the P1 prophage occurs bidirectionally [104]. Unique from *E. coli*, DnaA is able to use ADP when replicating P1 DNA enabling the preferential replication of the P1 chromosome [105]. The initiator protein RepA has multiple 19 bp binding sites across the P1 genome with one site, *incC*, being part of the replication origin (*oriR*) [96]. Additionally, *incA* is a regulatory locus for RepA. The RepA promoter of P1 is nested among *incC* iterons which suggests it is subjected to autorepression by the RepA protein [106]. Research has suggested that both *incC* and *incA* serve as regulatory elements of P1 DNA replication. For example, sequestering of RepA by *incA* is suggested to inhibit RepA accumulation at the origin of replication [107]. Additionally, *incC* iterons are able to bind RepA to their sites, thus limiting the amount of RepA [108]. Further, RepA forms a dimer which inactivates the RepA dimer and requires a bacterial chaperone to convert the dimer into an active monomer [109]. The partition module is downstream of *repA* and the *incA* introns, and is an operon composed of two genes, *parA* and *parB* which end at a centromere analogue *parA* [110].

The replication of the P1 genome is aided by the Cre recombinase located near its site of action, *lox*, and far from the DNA replication genes [96]. The *lox* site is unique in the P1 genome as only 1 site exists within the P1 genome [96]. The strongest promoter of *cre* is 500 bp upstream of *cre* and is overlapped by the C1 operator. This suggests that the *cre* promoter is inactive in P1 prophage [96]. The Cre protein is largely expressed

during lytic development and during infection where the Cre recombinase cyclizes the DNA of phages that harbor the *lox* site on each end [111]. Only a fraction of progeny phage will have *lox* site at each end of the chromosome suggesting only a minority of infectious virions use Cre to cyclize leaving the remaining infectious virions to rely on homologous recombination for cyclization [96].

The P1 lysogen ensures that daughter cells that do not inherit a P1 daughter chromosome die. The addiction operon of P1 is composed of genes *phd* and *doc*. This operon encodes a stable toxin protein, Doc, and an unstable antitoxin protein, Phd [96]. It is suggested that Doc reversibly blocks protein synthesis [96]. While P1 is maintained as a lysogen, Doc and Phd proteins form a complex leading to the repression of the auto regulated *phd doc* promoter [112]. In the case that P1 is not inherited by a daughter cell, Phd concentration will decrease allowing for Doc to suppress protein synthesis [113]. It has been suggested that the *phd doc* system of P1 is part of a larger system of post-segregationally cell killing in prokaryotes and shares a common mechanism with eukaryotic nonsense-mediated RNA decay [114].

P1 harbors subunits of the type III restriction-modification enzyme, EcoPI, as *res* and *mod*. The P1 Mod protein recognizes the DNA sequence 5'-AGACC and methylates the central adenine residue at the N-6 position [115]. The Res subunit exhibits double-strand cleavage nearly 25 bp to the 3' side of the recognition sequence and exhibits enzymatic cleavage only when in complex with Mod at a Res:Mod stoichiometry of 2:2 [116].

SimC is understood to encode for a superinfection exclusion protein as that gene is able to protect cells from superinfection by P1, P1 vir, and phage P7 [117]. *SimC* is part of the *simABC* operon with the genes encoding a precursor of periplasmic proteins [96]. The SimC protein is suggested to act on injecting phage and blocks injection into the host cytoplasm as SimC [96]. Similar to SieA of P22 and Imm protein of T4, SimC is suggested to act in the periplasm or in the cytoplasmic membrane to destroy the injected DNA or by preventing DNA access to the cytoplasm [96]. SimC is processed and requires SecA [118]. Cells carrying a low-copy number of sim^+ plasmids have a Sim^- phenotype [119]. This suggests that SimC is not active when P1 is a lysogen and only functions when P1 is undergoing its lytic cycle. This is due to dosage of *simC* being higher in the lytic cycle excluding superinfection during this process [96]. In this way, SimC function in a similar manner to T4 proteins ImmT and Sp as a means of inhibiting superinfection when undergoing the lytic cycle [120].

Unlike hallmark phages Mu and T4, phage P1 does not have a distinct intermediate stage of transcription [121]. The C1 operator sequences act as a switch for early genes while Lpa operating sequences act as a switch for late genes as RNA polymerase transcription initiates upon binding of Lpa [96]. Whole genome analysis of P1 reveals 17 operons to have C1 operators where σ^{70} promoters sequences are located [96]. These 17 promoters control 49 genes which are depressed in early lytic development [96]. Transcription among the early promoters of the lytic cycle are not initiated uniformly due to differences in affinity for the C1 repressor influenced by Lxc which modulates C1-operator binding [96]. Genes responsible for lytic replication are

transcribed within five minutes of induction and increased amounts of P1 DNA is detected nearly 15 minutes after induction [96]. This is nearly the time frame that cleavage of the packaging site in P1 DNA occurs [122]**.** Genes involved in the early lytic cycle do not remain highly active during the late lytic cycle as *E. coli* RNA polymerase-associated protein SspA attenuates transcription of early lytic genes [123]. The likely explanation to this is protein(s) transcribed during the late lytic cycle inhibit binding of RNA polymerase to early lytic genes.

P1 initiates DNA replication during its lytic cycle at *oriL* in a bidirection manner in the theta mode [96]. Later, the DNA replication program switches to rolling-circle replication [124]. While it is unclear where rolling circle replication initiates, there are several possible locations such as the pac site and the *rlf* operon which contains multiple sets of GATC sequences, a requirement for DnaA, and several IHF binding sites [96]. P1 encodes homologs of some *E. coli* replication protein such as DnaB helicase, single-stranded DNA-binding protein, Dam methyltransferase, and the theta subunit of DNA polymerase III, encoded by genes *ban, ssb, dmt,* and *hot*, respectively [96]. These genes are in separate operons which are all controlled by the C1 repressor [96]. Additionally, P1 lytic replication is understood to require host proteins DnaG, DnaC, DnaE, and DnaX which function as DNA primase, helicase loading factor, and the DNA polymerase III holoenzyme, respectively [96]. Antisense RNA transcribed from *kilA* may serve as a regulating mechanism of P1 linear DNA replication as the activity of *kilA* transcript interferes with transcription of *repL*. *RepL* is down-regulates P1 DNA replication during infection, and in late infection when P1 DNA concatemers accumulate [96]. P1 is able to

enhance recombination by use of gene *ref* which stimulates RecA-dependent recombination [96].

Late promoters of P1 resemble promotors of *E. coli* with the exception of the -35 hexamer [121]. Instead of the -35 hexamer, P1 promoters are identified as having a conserved 9 bp inverted repeat at position -22 which interacts with Lpa [125]. P1 late promoters require Lpa as well as the *E. coli* protein SspA [96]. Eleven late promoters control nearly 14 P1 operons [96]. Late gene products are synthesized at about 30 minutes after thermal induction of temperature sensitive P1 lysogens [96]. The late genes of P1 are involved with DNA replication, progeny morphogenesis, structural proteins, DNA packaging, and cell lysis [96].

P1 viral morphogenesis is similar to myophages in that the assembly pathway of heads and tails are independent [93]. P1 undergoes headful packaging which is initiated at the *pac* site [126]. Infectious P1 virions contain cyclically permuted, linear, double-stranded DNA with a terminal redundancy of nearly 10 kb [127]. The P1 virions contains substantially less structural proteins when compared to T4 [128]. For example, P1 lacks capsid decorating proteins present in phage T4 [96]. It is possible that the assembly pathway of the P1 capsid is similar to T4 in that a scaffold is required which is later released from the capsid allowing for the prohead to mature [96]. The Dar system is a dispensable system for P1 head morphogenesis, but leaves the P1 genome vulnerable to host restriction enzymes during infection. Additionally, the Dar system is thought to be packaged into the P1 capsid in a stepwise manner with the first two components being DarA and Hdf and the last component added to the P1 capsid being DarB [129]. Wild

type P1 has different head sizes [87], but the control of head size formation is not understood. The Dar system has been observed to impact head size when DarA and Hdf proteins are not present [129]. P1 stoichiometry of normal (T=13) capsid to small (T=7) capsids is 80 % and 20%, respectively, in wild type P1, but switches to 20% and 80% in *darA* and *hdf* single knockouts [129]. Since *darA* knockouts lack Hdf in the capsid, and *hdf* knockouts lack DarA it is assumed that both genes work in concert to guide capsid assembly and incorporation of the Dar system to the P1 virion [129]. Additionally, mutations in *prt* result in the formation of empty heads and non-processed DarA [130] while still having complete formation of tails [96]. The genes *pro* and *prt* compose a single transcriptional unit and encode the portal protein and the prohead protease, respectively [96].

Once the procapsid is formed, it can be bound to PacA. PacA forms a specific lesion within the *pacA* gene, termed *pac* site, within the concatermer produced by rolling circle replication and engages with the procapsid to package the P1 genome [131]. The initial headful packaging event starts at the lesion within the *pacA* gene, continues toward the site-specific recombination site *lox* which is nearly 4 kb away, and isn't completed until it packages the *lox* site again with an additional 4 kb of DNA [96]. This first headful packaging event contains two *lox* sites which are able to circularize during infection by homologous recombination or site-specific recombination [96]. The processivity of the PacA terminase appears to be three to four headfuls [131]. Cleavages within the *pac* locus are spaced apart by 13 bp and are sites of viral and bacterial methyltransferases which prohibit cleavage [122] [96]. Similar to T4, once the P1 Pac

terminase is bound to the double-stranded DNA within the *pac* locus, it is able to bind to the portal protein of P1 and package DNA into the procapsid in an ATP-dependent manner [132].

The P1 tail consist of a long cylindrical tube surrounded by a contractile sheath and originating from a baseplate with six tail fibers attached [96]. The head and tail assembly pathways are independent as mutations that affect tail morphogenesis do not have an effect on capsid formation [93]. P1 tail genes have homology to tail proteins of T4 and T4-like phages indicating a common ancestry of P1 and T4 tail functions [96]. There is a clear difference in the appearance of the baseplate with P1's being thinner and smaller when compared to the baseplate of T4 [93]. While the apparent differences between P1 and T4 are evident, there has not been a structural comparison between the two phages. Despite the prominent use of phage P1 in mapping the *E. coli* genome, the assembly process of P1 is not understood to the same detail as phage T4. The work presented here furthers the understanding of phage P1 assembly. In addition, this work investigates key motifs of DarB and reveals the structural composition of the P1 baseplate.

CHAPTER II

NEW INSIGHTS INTO THE STRUCTURE AND ASSEMBLY OF BACTERIOPHAGE P1

Introduction

Enterobacterial phage P1 is one of the earliest described temperate phages [82]. P1 is a well-known workhorse of molecular genetics due to its ability to transduce host DNA at high frequency [84] and across diverse bacterial species [85]. Phage P1 is now recognized as a member of a larger group of temperate phages that maintain themselves as plasmids and are associated with antibiotic resistance determinants [133] and virulence factors [134] in *E. coli*. Phage P1 and its relatives appear to be a unique phylogenetic group, with no clear relationships to other phages and are currently designated by the ICTV as the unclassified phage genus Punavirus.

The P1 genome encodes 119 genes, with 14 annotated structural proteins based mainly on an analysis of structural defects in P1 amber mutants [93,96]. However, a number of genes in P1's 94.5 kb genome have functions which are vaguely, if at all, defined. For instance, 18 genes are classified as 'putative' and 22 of the 119 predicted genes of bacteriophage P1 have no known function. Thus, it is still relevant to apply modern molecular methods to understand the structure of P1 to the level of other paradigm phages such as T4 and Lambda [33] [135]. Here, we present a complete proteomic profile of the P1 virion and identify essential structural proteins by a combination of whole virion mass spectrometry, transmission electron microscopy of P1

single-gene knockouts, and bioinformatic analyses. Additionally, we identify the peptide maturation cleavage site of the P1 major capsid protein.

Materials and Methods

Bacterial Strains and Phages

The bacterial strains and the parental phage P1CMclr100 (hereafter referred to as P1) used in this study were obtained from the Coli Genetic Stock Center, Yale University, or from our previous studies [129]. Unless otherwise noted, *E. coli* strains were cultured in LB (10 g/L Bacto Tryptone (BD Biosciences), 5 g/L Bacto yeast extract (BD Biosciences), 10 g/L NaCl (Avantor)) or LB agar (LB amended with 15 g/L Bacto agar) and incubated at 37 °C. P1 lysogens were cultured and maintained at 30 °C on LB amended with 10 ug/mL chloramphenicol (LB Cm-R) or 10 ug/mL chloramphenicol plus 30 ug/mL kanamycin (LB Cm-R Kan-R).

Production of Phage Lysates

Phage lysates were produced by thermal induction of P1 lysogens as described previously [129]. Briefly, an *E. coli* P1 lysogen was grown at 30 °C in LB Cm-R to an OD_{550} between 0.5 and 0.6. Next, the P1 lysogen was thermally induced by shifting the culture to 42°C [94] until the OD_{550} fell below 0.2. Then, chloroform was added (0.1% v/v) to the lysate. Finally, the lysate was centrifuged at 10,000 x g for 30 minutes, and the supernatant was stored at 4 °C until further use.

Transmission Electron Microscopy

Phages were stained with 2% uranyl acetate and imaged in a JEOL 1200 EX transmission microscope under 100 kV accelerating voltage as previously described [95,136].

pBAD24g plasmid construction

This method was reported first in another dissertation and is reported again here for clarity [137]. The gentamicin resistance cassette was PCR amplified from plasmid pMB838 with primers designed to have AatII and SacI restriction sites on the 5' and 3' end of forward and reverse primers, respectively. Similarly, the backbone of plasmid pBAD24 was PCR amplified with primers designed to have SacI and AatII restriction sites on the 5' and 3' end of forward and reverse primers, respectively. All PCR reactions were conducted using Phusion Hi-Fidelity PCR Master Mix (New England Biolabs) following the manufacturer's recommended protocol. The gentamicin resistance cassette was ligated into pBAD24 backbone using standard molecular biology techniques. P1 genes pacA, pmgA, pmgB, pmgC, pmgG and pmgR were cloned into pBAD24g (g for gentamicin resistance). These genes were PCR amplified using primers with XbaI and HindIII restriction sites on 5' end and were ligated into the respective restriction sites in pBAD24g following standard molecular biology techniques.

Generation of single-gene knockout mutants

This method was reported first in another dissertation and is reported again here for clarity [137]. Lambda Red recombinase mediated homologous recombination was

used to generate isogenic single-gene knock-out P1 mutants as described previously [129,138]. P1 was lysogenized into BW25113(pKD46) and colonies resistant to both chloramphenicol and ampicillin were selected. BW25113(pKD46) lysogenized with P1 and grown to an OD_{550} 0.5 in LB amended with 10 ug/mL chloramphenicol and 100ug/mL ampicillin. At this point, the culture was harvested, and made electrocompetent. Cells were transformed with linear PCR products containing a kan cassette and FRT sites with flanking regions homologous to the targeted insertion site as described previously [129]. The transformed cells were recovered in LB at 30 °C for 2 hours. After overnight incubation at 30 °C, chloramphenicol and kanamycin resistant colonies were selected and mutation was confirmed by PCR and sequencing.

Liquid Chromatography Tandem Mass-Spectrometry of Purified P1 Virions

Samples were analyzed by LC-MS/MS using an Ultimate 3000 nano-LC system (Thermo) coupled to an Orbitrap Fusion tribrid mass spectrometer (Thermo). 1 µl of sample was injected onto and separated by a 150 × 0.075 mm C18 column (Thermo Scientific Acclaim PepMap RSLC, 2 µm particle size) at a flow rate of 0.400 µl/minute. The total duration of the method was 60 minutes, with the gradient as follows: equilibration at 2% B (98% acetonitrile, 2% water, 0.1% formic acid), ramp to 45% B at 37 minutes, ramp to 90% B at 40 minutes and hold until 46 minutes, ramp down to 2% B at 47 minutes and hold at 2% B until the end of the run at 60 minutes. Eluent was introduced into the Fusion mass spectrometer by nano-ESI at a static voltage of 2450 V, with a transfer capillary temperature of 275 °C. Mass spectrometry data were acquired in positive mode at a resolution of 120,000 (at m/z 200) in the m/z range 400-1600. The RF

lens was set to 60%. Maximum injection time was 100 ms. Scans were acquired in top speed mode with a cycle duration (survey scan plus as many dependent scans as possible) set to 3 seconds. Monoisotopic peak determination was used in peptide mode. The intensity threshold for precursors of interest was 5.0*10^3. Charge states 1-6 were considered. Dynamic exclusion was set to 60 seconds with a mass tolerance of 10 ppm. MS/MS data were acquired by HCD at a fixed collision energy of 28% with a precursor ion isolation window of 1.6 m/z; fragments were detected in the ion trap at a rapid scan rate. Raw data were analyzed using Proteome Discoverer 2.1 (Thermo Scientific). MS/MS data were automatically extracted and matched against the P1 phage proteome database using the Sequest HT search engine. The generated list of detected proteins was further filtered to only include confident hits with a minimum of two unique peptides detected.

Bioinformatic Analyses

Protein sequences were searched against public databases by BLASTp [139]and HHpred [140] as appropriate. BLASTp searches were conducted against the NCBI nr database. HHpred searches were conducted against the PDB CIF70 (12 Oct 2021) and PFAM-A v35 databases at default settings. Results are reported at the matching database entry (as PDB or PFAM accession number) and the HHSearch percent probability score. In general, only matches with scores greater than 80% were considered credible, with most matches reported in this work having scores >95%.

Purification of Virion Sub-assembly by CsCl Step-Gradient Centrifugation

P1 mutants were induced in 100 mL LB as described above. 37 grams of ammonium persulfate (Millipore Sigma) was added to the lysate as it was shaking in ¼ increments every 15 minutes in a cold room. After the last addition, the lysate was allowed to shake for 1 hour. The lysate was centrifuged at 8,000 x g for 30 minutes at 4°C. The pellet was soaked in 5 ml SM buffer overnight. The lysate was filter sterilized using a 0.22 µM filter (Millipore Sigma). A shelf gradient of 1 ml 1.6 g/cc, 1 ml 1.4 g/cc, and 1 ml 1.2 g/cc CsCl, and 2 ml 20% sucrose was made in an Ultra-Clear™ centrifuge tube [9/16 X 3.5 inch (Beckman Coulter)]. The lysate was added to the top of the step gradient without disturbing the gradient. SM buffer was added using a serological pipette such that the tension of the SM buffer was flush with the opening of the tube. The tubes were loaded into a Beckman SW41 rotor and spun at 42,000 rpm for 2 hours at room temperature. The bands were extracted with an 18-gauge needle and dialyzed as described above. Phages were stored at 4°C for further use.

Complementation of P1 Δpmg Mutants

P1 knockouts were complemented in trans by induction of the mutant lysogen in a strain containing the expression vector of the complemented gene. Complementing genes were amplified by PCR from a P1 DNA template and cloned into pBAD24g at its XbaI and HindIII sites using standard molecular biology techniques [141]. Ligation products were transformed into competent *E. coli* 5-alpha cells (New England Biolabs) and selected by plating on LB agar amended with 20 µg/mL gentamicin. The plasmids were extracted as described above and were verified by sequencing before transforming

the relevant *E. coli* Δ*pmg* lysogen. Complemented P1 virions were generated by inducing the P1Δpmg lysogen in conjunction with inducing the corresponding ppmg complementing vector as follows. First, the transformed mutant lysogen was grown in LB at 30°C to an OD_{550} between 0.3 and 0.5. Next, L-arabinose (Millipore Sigma) was added to the culture to a final concentration of 1 mM and the incubation temperature was changed to 42°C until the OD_{550} dropped below 0.2. Finally, 0.1% chloroform was added to the lysate, the lysate was centrifuged at 10,000 x g at 4°C for 30 minutes, and the supernatant was collected. The phage lysate was used to infect the non-restricting host *E. coli* WA921 by combining 100 µL of WA921 at OD550 between 0.3 and 0.6 with 100 µL of the phage lysate with 5 mM $CaCl_2$ and allowed to incubate at room temperature for 30 minutes. The infection mixture was centrifuged for 1 minute at 13,000rpm and resuspended in 100 µL LB. The resuspension was plated on LB agar plates containing chloramphenicol and incubated at 30°C overnight.

Lysogeny Establishment Assay

Phage lysate was used to infect WA921 by combining 100 µL of bacteria at OD_{550} 0.3-0.5 with 100 µL of the phage lysate with a final concentration of 5 mM $CaCl_2$. The phage and bacterial mixture were allowed to incubate at room temperature for 30 minutes. The infection mixture was centrifuged for 1 minute at 13,000*rpm, the supernatant discarded, and the bacterial pellet was resuspended in 100 µL LB. The resuspension was plated on LB agar plates containing 10 ug/mL chloramphenicol and incubated at 30°C overnight.

Edman Analysis

Parental P1 was produced and purified by cesium chloride isopycnic centrifugation as described above. Purified P1 virions were run on an SDS-PAGE gel and the band corresponding to the mature major capsid protein was prepared according to the protocol described by the UC Davis proteomics core for Edman sequencing analysis (Edman Sequencing Analysis Molecular Structural Facility (ucdavis.edu)).

Results

Mass-Spectrometry Reveals the Proteome of the P1 Virion

Mass spectrometry is a sensitive application that can be used in a "bottom-up" approach to identify proteins within a sample [142]. This application has been used to identify structural proteins of multiple phages such as Pseudomonas aeruginosa phage O4 [142], Salmonella phage P22 [138], Listeria monocytogenes phage PSA [143], and coliphage T1 [144]. Although P1 is of significant historical importance, such modern approaches have not been applied to the P1 proteome. LC-MS/MS was conducted on trypsin digests of isopycnic CsCl gradient-purified P1 virions. DdrA is a known P1 structural protein [129] and yielded two unique peptides in this LC-MS/MS analysis, thus the value of two unique peptides was used as the cutoff for identifying a protein as a structural candidate.

Of the 119 open reading frames identified in P1 [96], 27 proteins meeting the two unique peptide cutoff were identified in our LC-MS/MS analysis of isopycnic CsCl purified P1 virions. Of these 27 proteins, nine of them produced a signal of two peptides

each. However, four of these proteins (TciA, Cre, Dmt, and Ant1) have well-defined non-structural roles in the P1 life cycle [96], thus they were excluded from further consideration as virion structural components. Thus, based on this analysis, 23 proteins are attributed to the morphogenesis of the P1 virion (Table 1), including the six proteins known to comprise the virion-associated Dar system [129]. This number is in general agreement with a previous study based on visualization of protein bands in SDS-PAGE that suggested 28 proteins make up the P1 virion with 15 composing the head, 9 the tail, and 4 identified to be in the head or tail [128].

The total number of structural proteins detected in P1 is somewhat less than T4, which is composed of 37 structural proteins [33] but more than bacteriophage lambda which is composed of 12 proteins [135]. Most of the identified structural proteins could be assigned a function and location in the virion based on previous studies [93,96] and additional bioinformatic analyses described below. Two novel proteins of previously unknown function, PmgC and PmgG, were also found to be virion-associated and are believed to play roles in tail or baseplate assembly, as discussed below. Much analysis of the P1 structure, particularly with respect to the tail and baseplate, will be discussed in the context of coliphage T4, a large myophage with extensive structural characterization [52] that also uses *E. coli* as its host.

Table 1 Structural proteins identified by LC-MS/MS of purified *E. coli* phage P1 virions. Predicted functions are assigned based on previous studies or new bioinformatic analyses. Proteins are listed in the order of their open reading frame (ORF) number in the P1 genome.

Detected Protein	Predicted Function	Molecular Mass (kDa)	Number of Unique Peptides	Coverage (%)	Location of Protein	ORF Number	Source	GenBank ID
Ulx	Aids DarB localization	17	3	20.5	Head	9	[129]	2777418
DarB	Inhibition of EcoK and EcoB endonucleases	251.4	37	25.8	Head	10	[129]	2777481
Prt	Portal protein	62.7	11	37.7	Head	11		2777482
Pro	Prohead protease	36.8	3	9.7	Head†	12		2777381
DdrB	Aids DarB addition to capsid	108.7	39	59.5	Head	22	[129]	2777413
DdrA	Antagonist of EcoA endonuclease	13	2	27.2	Head	23	[129]	2777414
DarA	Head size determinant and Dar system incorporation	69.4	27	67.4	Head	24	[129]	2777415
Hdf	Head size determinant and Dar system incorporation	22.1	7	39.9	Head	25	[129]	2777469
S	Tail fiber specificity	104.8	18	32.8	Baseplate	33	[96]	2777437
R	Baseplate wedge	15.9	3	26.3	Baseplate	34	This work	2777404

Table 1 Continued

Detected Protein	Predicted Function	Molecular Mass (kDa)	Number of Unique Peptides	Coverage (%)	Location of Protein	ORF Number	Source	GenBank ID
gp16	Baseplate structure*	31.3	2	11.5	Baseplate† *	35	This work	2777405
BplA	Baseplate wedge	53.5	7	21.5	Baseplate	36	This work	2777406
Sit	Tape measure protein	120.6	23	27.8	Tail & Baseplate	38	This work	2777408
Tub	Tail tube[a]	22.3	6	44.8	Tail	40	This work	2777479
PmgC	Tail adaptor	31.9	4	31	Tail	41	This work	2777480
BplB	Tail tube[a]	18.7	7	33.1	Tail	54	This work	2777444
PmgG	Baseplate-tail tube junction	20.5	2	20.2	Baseplate	55	This work	2777384
gp22	Tail sheath protein	56.9	22	56.1	Tail	57	[96]	2777386
gp23	Major capsid protein	62.2	24	47.5	Head	58	[96]	2777387

***Protein met the two-peptide threshold, but bioinformatic analysis does not support a clear structural role (see text and Table 3), therefore a putative role is assigned based on the annotation by Lobocka et al. [96].**

†These proteins may be incorporated into of the virion structure or serve as co-factors for virion assembly.

[a]These proteins can both putatively function as tail tube based on bioinformatic analysis (see text and Table 3).

P1 Head Proteins Identified by LC-MS/MS

Nine proteins were assigned to the P1 head, including the major capsid gp23, portal (Prt), and the six proteins that comprise the P1 Dar system (Table 1). A capsid scaffolding protein has not been identified in P1, and it is possible that the scaffold of P1 is the N-terminal portion of the major capsid precursor protein as observed in phages such as HK97 [145] or T5 [146]. A weak signal (3 peptides, 9.7% coverage) of the predicted prohead protease was identified in the P1 virion. This finding has also been observed in the proteomic analysis of phage lambda [135], thus it is not clear if this plays a structural role or simply represents incomplete removal of this protein during morphogenesis. It appears that P1 also lacks a decoration protein(s), stabilization, or vertex proteins as seen in phages such as T4 [147].

Proteolytic processing of the major capsid protein is a common feature of capsid formation in phages [148]. The transition between prohead I and prohead II of HK97 is characterized by the cleavage of a 104-residue polypeptide from the N-terminus of the major capsid protein by the phage-encoded protease gp4 [149]. The procapsid of phage T4 also undergoes proteolytic cleavage by the T4 protein gp21 [33]. In P1, processing of the major capsid protein gp23 has been presumed based on the discrepancy between the predicted length of the capsid gene product and the observed size of the mature P1 gp23 in SDS-PAGE gels. However, the processing site had not been determined. Edman degradation was employed to determine the primary sequence of the N-terminus of

mature P1 gp23. The N-terminal residues of the mature P1 gp23 were determined to be Ser-Val-Ala-Ala-Glu-Met (Figure 8), indicating 120 residues from the N-terminus of gp23 are removed during capsid morphogenesis. The predicted P1 prohead protease Pro is the most likely candidate to carry out this processing [96]. DarA is also believed to undergo proteolytic processing [130], however a similar preparation of the mature DarA protein did not produce a clear consensus signal from N-terminal sequencing.

The capsid protein of HK97 is considered as a model in understanding the ancestry of myophage capsid evolution [89]. There may be parallels between the capsid assembly of P1 and HK97 as a two-sequence alignment in HHpred shows the secondary structure of P1 Pro and HK97 gp4 are conserved between amino acids 89 and 198 of the P1 sequence. It has been shown that linker insertion mutations in residues 130, 140, and 179 in gp4 of HK97 disrupt the activity of gp4 [89], and the conservation of this region in P1 would suggest a similar phenotype. HHpred alignment between the HK97 capsid protein gp5 and P1 gp23 revealed conservation in secondary structure between gp23 residues 307-469, respectively. This region includes the C-terminal portion of the gp5 E-loop and the betaC through betaF domains but lacks loop and alpha4. This suggests that HK97 may serve as a model for P1 head maturation.

A

Cycle	Amino Acid	Pico Mole
1	Ser	27
2	Val	30
3	Ala	23
4	Ala	-
5	Glu	13
6	Met	7.7

B

Figure 8 Edman degradation of P1 major capsid protein gp23. (A) Edman degradation reveals the first six amino acids of the N-terminus of gp23 purified from P1 virions. (B) The amino acids identified in the open reading frame of gp23, in which the first 120 residues are processed from the N-terminus of gp23.

P1 Tail Proteins Identified by LC-MS/MS

As a myophage, the P1 tail is expected to be composed mainly of tail tube and sheath protein. The P1 tail sheath protein has been identified as gp22 based on previous work [96] and this is confirmed here with 51.6% peptide coverage (Table 1) and a strong HHpred match to the T4 tail sheath gp18 (Table 3). The identity of the tail tube protein has been more difficult to confirm. Both Tub and BplB produced similar signals in the LC-MS/MS analysis (Table 1) and both have strong matches (>99.5%) to the T4 tail tube protein gp19 (5IV5_IB) as determined by HHpred. BLASTp alignment of Tub and BplB returns a short region of amino acid similarity (22 residues, 30% identity, E=0.015) that suggests these proteins may be the result of a gene duplication event in the distant past. Alignment of these two proteins in HHpred shows strong secondary sequence conservation, and their similar relationships to T4 gp19 indicates that both proteins could plausibly function as the major tail tube component.

There has been a historical debate on the identity of the P1 tape measure protein, with both gp6 and Sit annotated with this function [96]. While gene 6 knockouts produced variable tail length in P1 [93], analysis by HHpred suggests it is a baseplate component, related to both the T4 gp27 central spike (P17172, 99%) and the phage Mu baseplate hub gp44 (1WRU_A, 99%). P1 gp6 is also too short to function as a tape measure protein: at 338 aa, gp6 could not determine the length of the 120 nm P1 tail based on the assumption of 1.5 angstroms length per residue in an alpha-helical protein [150]. At 1,140 aa in length, Sit is a more appropriate size for the tail tape measure protein. Protein Sit contains a recognizable soluble lytic transglycosylase (SLT) domain (HHpred 5OHU_A, 99%) which spans residues ~750-900 of the Sit protein [151], and such SLT domains are a feature of some phage tail tape measure proteins [128,144]. Analysis of Sit for predicted secondary structure by Quick2D [140] shows that the N-terminal two thirds of the protein are pre-dominantly alpha-helical. This is consistent with tape measure protein structure, and, in addition, the length of the Sit alpha-helical region is nearly 750 residues which is congruent with the length of the 120 nm P1 tail [152].

Other P1 proteins detected by LC-MS/MS (Table 1) and identified by HHpred (Table 3) to have a predicted function are gp24 and PmgC. The total number of proteins identified to be part of tail assembly in P1 is similar to the number required for tail assembly in T4 [33]. HHpred analysis showed P1 protein gp24 to have strong similarity (94.6% probability) to the T4 tail terminator protein gp15 (Table 3), indicating P1 gp24 is located at the top of the tail at the tail-head interface [51]. This analysis also identified

PmgC as a possible head-tail joining component, with matches to the T7 gp11 tail adaptor. Direct two-sequence alignments in HHpred showed alignment (84% probability) of P1 PmgC to the T4 gp13 neck protein [51], indicating a role in head-tail connection.

P1 Baseplate Proteins Identified by LC-MS/MS

The myophage baseplate is the nucleating site for tail assembly and is therefore assembled before the tail [33]. In the well-studied myophage T4, nearly 150 subunits form oligomeric components which compose the six wedges that make up the T4 baseplate [33]. The wedges form around a central hub with the aid of trimeric proteins gp9 [153] and gp12 [154]. There are 15 proteins identified in the T4 baseplate including the tape measure protein [52]. While the assembly process of the P1 baseplate is not known, mass-spectrometry identified five proteins which are predicted to play a role in baseplate structure and have been assigned roles by bioinformatic analysis: S, R, BplA, PmgG, and gp6 (Table 1, Table 3). The tape measure protein Sit, which is a component of both the baseplate and tail, is discussed below. P1 proteins R and BplA are baseplate wedge components orthologous to T4 gp8 and gp6, respectively, based on HHpred analysis (Table 3). P1 gp6 is annotated as the baseplate hub orthologous to T4 gp27 as noted above, and PmgG is here annotated as the T4 gp48-like baseplate/tail junction. P1 S is the well-characterized tail fiber and specificity determinant responsible for host recognition [96].

Identification of additional morphogenetic genes in P1

A previously published analysis of the P1 genome [96] annotated the functions of 22 proteins as pmg (putative morphogenetic function, 16 genes) or upf (unknown protein function, 6 genes). To better understand the role of these P1 genes, a collection of 29 single-gene knockouts were produced. Of these, 24 mutants had no observable plaquing defect and did not exhibit changes in antirestriction activity. However, five of these single-gene knockouts (*pmgA*, *pmgB*, *pmgC*, *pmgG*, and *pmgR*) were unable to produce infectious virions under normal induction and plating conditions, indicating that these genes encode essential functions.

To confirm that defects in plaque formation were due to each specific gene deletion, we complemented each pmg knockout lysogen by cloning the cognate gene into pBAD24g and expressing the vector in trans; this complementation vector was also used as to transform the *E. coli* host used in soft agar overlays. While plaque formation was rescued for P1Δ*pmgR* by this approach, plaque formation was not restored for the remaining four mutants. Phage plaque formation in soft agar overlays requires multiple cycles of phage infection, replication, and lysis to produce a visible plaque [155]; It is possible that gene dosage or expression issues of the in trans system reduced phage burst sizes and thus prevented the recovery of the plaquing phenotype. To confirm if complemented P1Δ*pmg* knockouts were able to produce infectious virions, we moved to an experimental approach that would measure lysogen establishment as a marker for successful P1Δ*pmg* complementation (Table 2).

Table 2 Complementation of putative morphogenetic gene knockouts.
Complementation of all mutants except for pmgR did not rescue the plaque forming phenotype. Thus, complementation was observed by the restoration of ability to produce infections virions that can establish lysogens in new hosts.

Lysate	Lysogens established/mL induced lysate	Fold Increase in lysogen establishment	Estimated Virions/mL in Induced Lysate
P1Δ*pmgA*	$1.8x10^1$	-	$6.0x10^2$
P1Δ*pmgA*+ppmgA	$2.0x10^2$	11	$6.7x10^3$
P1Δ*pmgB*	$1.5x10^1$	-	$5.0x10^2$
P1Δ*pmgB*+ppmgB	$7.0x10^1$	4.7	$2.3x10^2$
P1Δ*pmgC*	$2.2x10^1$	-	$7.3x10^2$
P1Δ*pmgC*+ppmgC	$1.3x10^2$	6.1	$4.4x10^3$
P1Δ*pmgG*	$3.4x10^2$	-	$1.1x10^4$
P1Δ*pmgG*+ppmgG	$1.6x10^5$	460	$5.3x10^6$
P1Δ*pmgR*	$1.3x10^1$	-	$4.3x10^2$
P1Δ*pmgR*+ppmgR	$2.4x10^3$	180	$8.0x10^4$
P1	$3.0x10^5$	-	$1.0x10^7$

Each P1Δ*pmg* lysogen was induced in the presence or absence of its cognate complementing plasmid, and the lysates used to infect 100 µL aliquots of *E. coli* WA921. Lysogen formation was measured by plating infected WA921 on LB agar with and without 10 µg/ml chloramphenicol and to enumerate colony forming units. By comparing the number of lysogens established between complemented P1Δ*pmg* lysates

and their respective uncomplemented P1Δ*pmg*, complementation of each P1Δ*pmg* background was measured. We believe this to be a sensitive assay as P1 virions establish lysogeny at a consistent probability of ~30% [156]. As shown in Table 2, lysates from all in trans-complemented P1Δ*pmg* mutants produced between 5 to nearly 500-fold more lysogens than the uncomplemented P1Δ*pmg* backgrounds, indicating the restoration of the parental phenotype with varying degrees of efficiency.

Assigning Roles to the Newly Identified Essential Genes in P1

In tailed phages, tail and head assembly pathways are independent with the only interaction between the pathways occurring when the completed head and tail combine to form a functional virion [33]. Previous work on P1 predicted the roles of multiple morphogenetic genes by examination of P1 amber mutant lysates by transmission electron microscopy (TEM) [93]. Given the locations of the essential pmg genes in the P1 genome, we hypothesized that these genes play roles in phage morphogenesis. Virions of the five P1 pmg deletions and a pacA deletion were purified by cesium chloride step gradients and examined by TEM (Figure 9). The pacA mutant disables the P1 DNA packaging motor complex [126] and was used as a control for purification of incomplete P1 virion components. In T4, defects in DNA packaging are known to result in phage heads stalled at an incomplete prohead state which lack packaged DNA and are unable to join with the assembled phage tails [157].

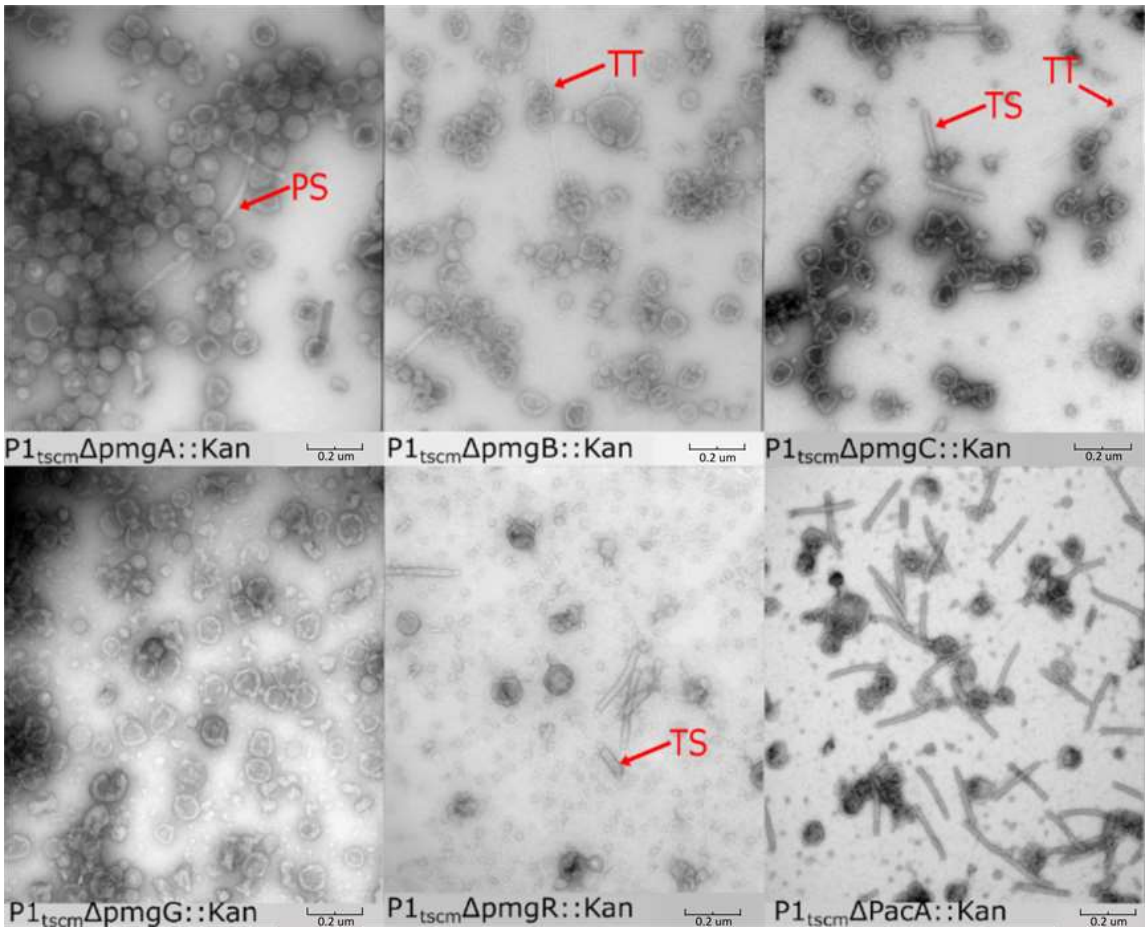

Figure 9 Transmission electron micrographs of CsCl step-gradient purified P1 morphogenic gene knockouts and parental P1. PS is polysheath, TT is tail tube, and TS is truncated sheath.

In the pacA knockout, phage proheads and complete phage tails were observed in roughly equal proportion (Figure 9), indicating that purification by cesium chloride step gradient can recover P1 virion components. Deletion of *pmgC* and *pmgR* produced lysates containing tails and incomplete heads but no complete P1 virions. This phenotype is consistent with the predicted function as a part of the head-tail connection complex described above. PmgR does not have a relationship detectable by HHpred to other proteins of known function, however the similar phenotype in the *pmgR* and *pmgC* deletions suggest a role in head completion or head-tail joining. In contrast, deletions in

pmgA, pmgB, and *pmgG* produced no observable complete tails, but did produce possible aberrant polytube and poly-sheath structures as have been reported previously [93]. The heads of *pmgA, pmgB,* and *pmgG* knockouts are amorphous and resemble those of P1 mutants where tail assembly was perturbed [93], suggesting that completed P1 heads are unstable in purification if they remain incomplete due to lack of tail attachment. Analysis of PmgA and PmgG by HHpred suggest roles as baseplate components, with PmgA similar to the T4 gp25 inner baseplate wedge (5IW9_B, 99.4%) and PmgG similar to the T4 gp48 baseplate-tail tube junction protein (5IV5_R, 97.3%). PmgB is not related to proteins of known function, but the location of the pmgB gene immediately upstream of sit, the P1 tape measure gene, suggests a possible role as a tape measure chaperone [152].

Taken together, the data suggests that PmgA, PmgB, and PmgG play essential roles in tail morphogenesis while PmgC and PmgR play roles in head completion or head-tail attachment. Only PmgC and PmgG were detected in the virion proteomic analysis (Table 1). Failure to detect PmgB would be consistent with its hypothesized role as a tail chaperone and would not be expected to appear in the assembled virion [152]. Lack of LC-MS/MS signal for PmgA and PmgR may be due to their small size (13.2kDa and 8.3kDa, respectively), low copy number in the virion, and low numbers of peptides produced on trypsin digestion. While PmgA is predicted to act as a baseplate component, PmgR has no predicted function, and its deletion produces a phenotype of unjoined heads and tails. PmgR may play a role in head completion or head-tail joining but cannot be conclusively determined to be part of the virion at this time.

Identification of Additional Structural Proteins

To identify additional structural components of the P1 virion, HHpred searches were conducted against all proteins of unknown function (annotated as pmg, upf, uhr or upl) and proteins implicated in morphogenesis in previous studies [96]. This analysis identified three additional P1 proteins, UpfC, gp5, and gp26, as having structural roles in P1 (Table 3). UpfC shows a high-quality alignment with the T4 gp5.4 spike tip, which is attached to the C-terminal tip of the gp5 cell-puncturing tail needle [52,158]. The deletion of upfC did not result in any noticeable disruption to P1's plaquing ability, indicating this gene is dispensable in P1. In T4, gp5.4 also appears to not be essential for plaque formation (Petr Leiman, personal communication). P1 gp5 aligned in HHpred with the T4 gp5 tail needle/lysozyme, however this alignment spanned only T4 gp5 residues 1-100 and 385-500, which corresponds roughly to the known N-terminal OB-fold domain and C-terminal beta-helix domain of T4 gp5, which forms the cell-puncturing needle structure [52,158]. This suggests that the P1 virion contains a cell-puncturing needle similar to that of T4, but this lacks the lysozyme activity associated with its T4 ortholog. Finally, P1 gp26 showed a good alignment with T4 gp53, which forms a part of the baseplate core in T4 [52] All of these proteins are relatively small (~10 – 20 kDa) and probably present in low copy number based on their T4 counterparts (UpfC, gp5 and gp26 present in 1, 3 and 6 copies per virion, respectively) [52]. This may be why these proteins were not identified in the LC-MS/MS analysis of P1 and illustrates the limitations of this approach.

Conclusion

Bacteriophage genomes are known for their modular arrangements, with genes related to a specific function colocalized in the genome [135,159]. While many of the genes of P1 are bundled in such modules, multiple genes depart from this arrangement (Figure 5). For example, P1 gene 23 encoding the major capsid protein is located between gene 22, encoding the tail sheath, and parS, encoding part of the P1 prophage partition system. The placement of P1 23 is a departure from the placement of the major capsid gene in other model phages such as T4, lambda, and HK97, where the major capsid gene is closely associated with other genes related to capsid morphogenesis such as the scaffold, portal or prohead protease [135,159,160]. Moreover, the 15 genes that have bioinformatic and LC-MS/MS evidence to support their role in tail or baseplate assembly (Figure 5) are scattered through the P1 genome. Genes numbered 30-41 are involved in baseplate, tail, and tail fiber morphogenesis and assembly, but genes bplB, gp22, gp6, and gp24 are in positions 54, 57, 83, and 84, respectively. The placements for these genes depart from lambda genome architecture [135], but somewhat resemble the gene placement of T4 [159]. Further, genes 88 through 103 have been implicated as "head structure and processing" [96], but our work suggests that is not the case, as deletion of many of these genes does not appear to prevent the phage from making infectious virions.

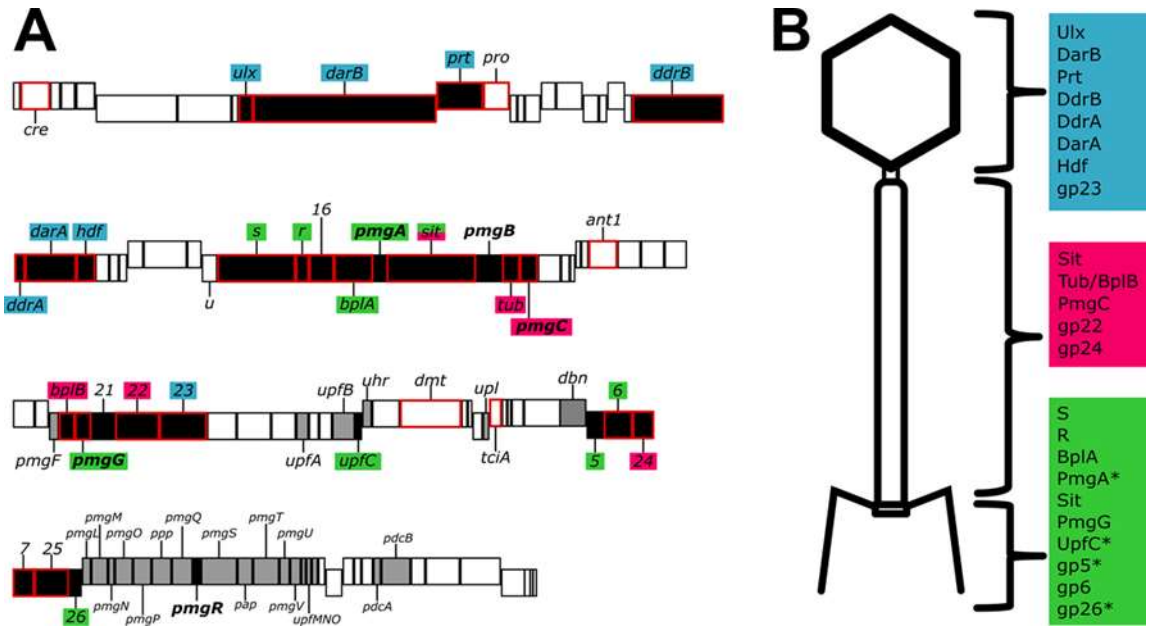

Figure 10 The P1 genome and composition of the virion. A) Genome map of phage P1 summarizing P1 morphogenesis genes. The size of the boxes indicates of the size of each ORF, with genes encoded on the plus and minus strands staggered up and down, respectively. Genes colored black are involved in phage morphogenesis, based on the work presented here or in previous studies [96]. Products of genes outlined in red were detected in LC-MS/MS above the two-peptide threshold. Genes with bold labels were determined to be essential for morphogenesis by construction of single-gene deletions. The deletion of gray genes produced no antirestriction phenotype and no apparent defect in plating, thus these genes are considered non-essential with unknown function. Gene labels are color coded to indicate their predicted location in the virion (panel B), with blue indicating the head, magenta the tail, and green the baseplate. Proteins with unknown functions are not colored. B) A schematic of the P1 virion, partitioned into three components of the head, tail, and baseplate (including tail fibers). Structural proteins are assigned to each area based on their predicted roles; color coding corresponds to the coding of gene labels in panel A. Sit is listed in both the tail and baseplate as its role as tape measure protein makes it a part of both components. Proteins PmgA, UpfC, gp5 and gp26 are marked with an asterisk as these are predicted to be part of the virion based on bioinformatics but were not detected by LC-MS/MS.

A comparison of the proteins involved in baseplate formation in P1 and the paradigm phage T4 shows some significant disparities in their compositions. The T4 baseplate proper (excluding side tail fibers) is reported to contain 15 proteins, eight of which are proposed to form a conserved baseplate core [52]. In contrast, our analysis has

been able to confidently assign only nine proteins to the P1 baseplate (Supplementary Table 5), suggesting the baseplate of P1 possesses a simpler structure than that of T4. Phage P1 has not been reported to possess short tail fibers analogous to T4 gp12, and such fibers are not evident in images showing P1 infection [161]. Thus, it is perhaps expected that orthologs of T4 gp10, gp11 and gp12 are absent in P1, as gp9 and gp10 are involved in the attachment of T4 gp12 to the baseplate [52]. Of the eight described core T4-like baseplate components, P1 possesses detectable orthologs of seven (Table 4). This accounts for nearly the complete conserved inner baseplate core, with the exception of a clear ortholog of T4 gp7, which makes intimate contacts with gp6 in a 2:1 (gp6:gp7) heterotrimer and is conserved in myophage baseplates [52,162]. Searches of all annotated P1 proteins in HHpred was also unable to identify an ortholog of T4 gp7. A second baseplate-tail tube adaptor protein found in T4, gp54, is absent in P1, suggesting PmgG could fulfill this role alone. Electron micrographs showing the P1 baseplate (e.g., [161]) show a structure that is considerably more compact than the baseplate of T4, which would be consistent with a simplified baseplate structure in P1 relative to that of T4.

A T4 gp5-like needle ortholog was identified in P1 as gp5, but the P1 version of this protein appears to lack the lysozyme domain present in T4 gp5. This activity is likely supplied in P1 by the SLT domain found near the C-terminus of the predicted P1 tape measure protein Sit. Five additional P1 proteins were identified as virion-associated with a weak LC-MS/MS signal (gp7, gp16 and gp25), or were shown to be essential for virion morphogenesis (PmgR, PmgB). Protein similarity searches including HHpred were unable to identify possible roles for these proteins. These proteins may fulfill some

of the functions associated with the "missing" T4 orthologs but are sufficiently diverged as to be no longer recognizable or be horizontal gene acquisitions. They may also play important non-structural roles in morphogenesis, such as acting as chaperones or cofactors analogous to proposed roles for T4 gp26 and gp51 [48].

In conclusion, while many of the essential genes of P1 have been elucidated using classical genetic approaches [93], approximately half of the genes in the P1 genome have no defined role. We have provided evidence that out of 29 genes with unknown function, five are essential for P1 morphogenesis while the remaining genes show no defect in plaquing, indicating that these functions are dispensable. In addition, proteomic analysis of the virion validated previously identified P1 structural genes and identified two additional genes as being incorporated into the virion (Table 1). Combining previously published work on P1 morphogenesis [93] and whole genome analysis [96] we are able to confidently assign locations to most of the P1 structural proteins identified by mass spectrometry. Taken together, this provides an updated genome annotation that can assign functions to multiple P1 proteins including baseplate, tail, and head-tail joining components (Figure 10). Updated bioinformatic analyses suggest that many P1 structural proteins have analogs in the well-studied myophage T4, including the capsid, portal, tail sheath, tail tube, nine baseplate components, and three tail completion or head-tail joining components. Unlike T4, P1 does not appear to encode a separate scaffolding system for head assembly and may rather rely on a scaffold domain located at the N terminus of the major head protein as seen in phages such as T5 and HK97 [52,145]. Also, unlike T4, P1 appears to use a tail-associated

muralytic enzyme that is fused to the end of the tape measure protein as seen in phages such as T5 [163] and TM4 [164] rather than a separate tail lysozyme analogous to T4 gp5 [147]. Remaining unresolved is the true identity of the P1 major tail tube subunit, as two candidate proteins, Tub and BplB, were identified as part of the virion and have matches to the T4 tail tube protein gp19. We have developed a more complete picture of P1 assembly by identifying additional essential proteins required for morphogenesis and the capsid protein processing site.

Table 3 Results of HHPred searches for P1 proteins identified as virion-associated by LC-MS/MS analysis or essential by genetic knockouts. Proteins Gp7, Gp25 and Gp16 did not return high-quality matches in HHPred and their functions remain unknown.

	P1 protein	**Match accession**	**Probability (%)**	**Function**	**PMID**
Proteins identified by LC-MS/MS					
	gp23	5VF3_L	100	Bacteriophage T4 isometric capsid	28893988
	Prt	3JA7_B	100	Portal protein gp20; VIRAL PROTEIN; 3.63A {Enterobacteria phage T4}	26144253
	Pro	5JBL_A	99.44	Prohead core protein protease; protease pentamer, phage T4, prohead, HYDROLASE; 1.943A {Enterobacteria phage T4}	27667692
	DdrB	ND		Antirestriction component	28509398
	DarA	ND		Antirestriction component	28509398
	DarB	ND		Antirestriction component	28509398

Table 3 Continued

	P1 protein	Match accession	Probability (%)	Function	PMID
Proteins identified by LC-MS/MS					
	Hdf	ND		Antirestriction component	28509398
	DdrA	ND		Antirestriction component	28509398
	Ulx	ND		Antirestriction component	28509398
	gp22	3J2M_Y	100	Tail sheath protein Gp18; bacteriophage T4, phage tail terminator protein, phage sheath protein, VIRAL PROTEIN; 15.0A {E	23434847
	Tub	5IV5_IB	99.94	Tail tube protein gp19; T4, baseplate-tail tube complex, pre-attachment, bacteriophage, bacterial virus, hexagonal, memb	27193680
	gp6	1WTH_D	99.06	Baseplate structural protein Gp27; Triple-stranded beta-helix, OB fold, pseudohexamer, T4 tail lysozyme, HUB, gp5-gp27,	15701513
	gp7	None >50%			
	gp25	None >50%			
	Sit	5OHU_A	98.92	Soluble lytic murein transglycosylase; Lytic Transglycosylase, LYASE; HET: PO4; 2.2A {Pseudomonas aeruginosa}	29632171
	BplA	5HX2_D	100	Baseplate wedge protein gp6; T4, baseplate, complex, VIRAL PROTEIN; 3.8A	26929357

Table 3 Continued

	P1 protein	Match accession	Probability (%)	Function	PMID
Proteins identified by LC-MS/MS					
	R	6HHK_C	98.72	Gp105; bacteriophage baseplate protein, VIRAL PROTEIN; HET: MSE; 2.38A {Listeria phage A511}	30606715
	gp24	4HUD_C	94.58	Tail connector protein Gp15; Bacteriophage T4, phage tail terminator protein, gp15, VIRAL PROTEIN; 2.7001A {Enterobacter	23434847
	gp16	None >50%			
	BplB	5IV5_IB	99.79	Tail tube protein gp19; T4, baseplate-tail tube complex, pre-attachment, bacteriophage, bacterial virus, hexagonal, memb	27193680
	PmgC	7BOX_N	87.4	T7 gp11 tail adaptor	32266588
	PmgG	5IV5_r	97.3	T4 tail tube assembly protein gp48; baseplate-tail tube junction	27193680
Additional proteins identified as essential	**P1 protein**	**Match accession**	**Probability (%)**	**Function**	**PMID**
	PmgA	5IW9_B	99.4%	T4 gp25 inner baseplate wedge, initiator of sheath polymerization	27193680
	PmgB	None >50%		Tape measure chaperone?	

Table 3 Continued

Additional proteins identified as essential	**P1 protein**	**Match accession**	**Probability (%)**	**Function**	**PMID**
	PmgR	None >50%			
Other proteins with bioinformatically assigned structural roles	P1 protein	Match accession	**Probability (%)**	**Function**	**PMID**
	UpfC	4KU0_D	**99.4%**	**T4 gp5.4 baseplate needle structure**	
	gp5	1WTH_A	**99.67%**	**T4 gp5 tail needle**	
	gp26	5HX2_F	**98.9%**	**T4 gp53 wegde component**	

Table 4 Table 4 Baseplate components of coliphage T4 compared to baseplate components identified in phage P1. Functions of P1 proteins were predicted using HHPred at default settings. Detection of the P1 protein by LC-MS/MS of purified P1 virions is also shown (see Table 1). Nine P1 proteins with orthologs to T4 baseplate components were identified, most of these are in the baseplate core. P1 proteins with no matches in HHPred are denoted with dashes. T4 proteins and their functions are adapted from Taylor et al. 2016 [165].

T4 protein		**T4 function**	**P1 ortholog**	**Structure/ domain**	**% Probability**	**Detected in LC-MS/MS**
Baseplate core proteins						
	gp5.4	Spike tip	UpfC	4KU0_D	99.4%	No
	gp5	Hub, needle, tail lysozyme	gp5	1WTH_A	99.67%	No
	gp6	Wedge	BplA	5HX2_D	100	Yes
	gp7	Wedge	--	--	--	--
	gp25	Wedge	PmgA	5IW9_B	99.4%	No
	gp27	Hub	gp6	1WTH_D	99.06	Yes
	gp48	Baseplate-tail tube junction	PmgG	5IV5_r	97.3	Yes
	gp53	Wedge	gp26	5HX2_F	98.9%	No

Table 4 Continued

T4 protein	T4 function	P1 ortholog	Structure/ domain	% Probability	Detected in LC-MS/MS
Baseplate other components					
gp8	Wedge	R	6HHK_C	98.72	Yes
gp29	Hub, tape measure	Sit	5OHU_A	98.92	Yes
gp54	Baseplate-tail tube junction	--	--	--	--
Baseplate tail fiber attachment					
gp9	Wedge, LTF attachment site		--	--	--

Table 4 Continued

T4 protein	T4 function	P1 ortholog	Structure/ domain	% Probability	Detected in LC-MS/MS
gp10	Wedge, STF attachment	--	--	--	--
gp12	Baseplate outer rim, STF	--	--	--	--

CHAPTER III

SITE-DIRECTED MUTAGENESIS OF DARB AND DARB CAPSID LOCALIZATION SIGNAL

Introduction

Bacteriophages are obligate parasites of bacteria. This interaction has led to strategies that bacteria utilize to keep the bacteriophage from successfully infection [68]. Thus, bacteriophage have developed avenues to evade bacterial defense systems [68]. Bacteriophage defense systems can be broadly characterized into three classes [68]. A bacterial defense system that has significance in the laboratory are restriction enzymes.

The study of bacteriophage P1 led to the discovery of restriction enzymes [166]. Now, restriction enzymes are well understood and are used routinely in the laboratory as a means for molecular cloning. Phage P1 was identified to overcome restriction enzymes by means of a defense against restriction genes [94]. Key genes in the defense against restriction phenotype are DarA and DarB [94]. Additionally, P1 is understood to package proteins involved in the anti-restriction phenotype in a step wise manner [129]. Internally packaged proteins have been identified in T4 and have been studied to the extent that a capsid localization signal has been identified in T4 internal protein III [36].

Phage T4 has served as a model for Myoviridae. Phage T4 packages multiple proteins to its' capsid notably the protein Ocr mimics DNA and acts as an inhibitor for bacterial restriction endonucleases [37]. The N-terminus of T4's IPIII has been used to packaged green fluorescent protein to the T4 capsid [132]. Further, MNase has been packaged to the T4 virion and has exhibited enzymatic activity while packaged in the T4 capsid [132].

In phage P1 it is unknown how the defense against restriction proteins are packaged to the P1 virion. To understand how the Dar system is localized to the P1 virion truncations in N-terminus of DarB were made and truncations mutants were expressed in P1 lysogens deficient in DarB expression. Further, the capsid localization signal was fused to mCherry and packaging into the P1 capsid was observed by fluorescent microscopy. Moreover, site-directed mutagenesis on DarB was conducted in helicase and methyltransferase domains of DarB.

Materials and Methods

Bacterial Strains and Phages

The bacterial strains and the parental phage P1CM*clr*100 (hereafter referred to as P1) used in this study were obtained from the Coli Genetic Stock Center, Yale University, or from our previous studies [129]. The full list of strains is shown in supplementary Table S1 and handled as previously described [129]. Unless otherwise noted, *E. coli* strains were cultured in LB (10 g/L Bacto Tryptone (BD Biosciences), 5 g/L Bacto yeast extract (BD Biosciences), 10 g/L NaCl (Avantor)) or LB agar (LB amended with 15 g/L Bacto agar) and incubated at 37 °C. P1 lysogens were cultured and maintained at 30 °C on LB amended with 10 μg/mL chloramphenicol (LB Cm-R) or 10 μg/mL chloramphenicol plus 30 μg/mL kanamycin (LB Cm-R Kan-R).

Table 5 Strains and plasmids used in this study.

Strains, Phages, or Plasmids	**Genotype or Relevant Characteristics**	**Reference/Source**
E. coli		
WA921	F- *thr*-1 *leuB*6(Am) fhuA21 *lacY*1 *or Δ(cod-lacI)*6 *glnX*44(AS) λ^- *metB*1 *thiE*1 *hsdS*3	CGSC
W3110	F- λ^- *IN(rrnD-rrnE)*1 *rph*-1	CGSC
WA2379	F- *leuB*6(Am) *fhuA*21 *lacY*1 *or Δ(cod-lacI)*6 *glnX*44(AS) λ^- *metB*1 *thiE*1 *deoB*20	CGSC
WA960	F- *leuB*6(Am) *fhuA*21 *lacY*1 *or Δ(cod-lacI)*6 *glnX*44(AS) λ^- *metB*1 *thiE*1	CGSC
5-alpha competent	Used for its high efficiency transformation	New England Biolabs
Phages		
P1CM*clr*100	*Clr*100 ts; Chloramphenicol resistant	CGSC
P1Δ*darB*	In-frame deletion of *darB* in P1CM*clr*100	[129]
Plasmids		
pBAD24	Empty cloning vector	Laboratory Stock

Table 5 Continued

Strains, Phages, or Plasmids	Genotype or Relevant Characteristics	Reference/Source
pdarBΔ1-4	P1 darB missing the first 4 codons cloned into pBAD24	[137]
pdarBΔ1-9	P1 darB missing the first 9 codons cloned into pBAD24	[137]
pdarBΔ1-14	P1 darB missing the first 14 codons cloned into pBAD24	[137]
pdarBΔ1-19	P1 darB missing the first 19 codons cloned into pBAD24	[137]
pdarBΔ6-30	P1 darB missing codons 6-30 cloned into pBAD24	This Work
pdarBΔ11-30	P1 darB missing codons 11-30 cloned into pBAD24	This Work
pdarBΔ16-30	P1 darB missing codons 16-30 cloned into pBAD24	This Work
pdarBΔ21-30	P1 darB missing codons 21-30 cloned into pBAD24	This Work
pFTSKi-tetR-mCherry	Template for PCR amplification of tetR-mCherry	[167]

Table 5 Continued

Strains, Phages, or Plasmids	Genotype or Relevant Characteristics	Reference/Source
pBAD24_N30-tetR-mCherry	mCherry fused to the first 30 amino acids of P1 darB cloned into pBAD24	[137]
pBAD24_tetR-mCherry	pBAD24 expressing tetR-mCherry alone	[137]
pBAD24_darBT93A	pBAD24 expressing DarB with a site-directed mutation at amino acid 93 changing the tyrosine to an alanine	This Work
pBAD24_darBT162A	pBAD24 expressing DarB with a site-directed mutation at amino acid 162 changing the tyrosine to an alanine	This Work
pBAD24_darBT1118A	pBAD24 expressing DarB with a site-directed mutation at amino acid 1512 changing the arginine to an alanine	This Work

Production of Phage Lysates

Phage lysates were produced by thermal induction of P1 lysogens as described previously [129]. Briefly, an *E. coli* P1 lysogen was grown at 30 °C in LB Cm-R to an OD550 between 0.5 and 0.6. Next, the P1 lysogen was thermally induced by shifting the culture to 42 °C [94] until the OD_{550} fell below 0.2. Then, chloroform was added (0.1% v/v) to the lysate. Finally, the lysate was centrifuged at 10,000 × g for 30 min, and the supernatant was stored at 4 °C until further use.

Generation of DarB-tetR-mCherry Capsid Localization Sequence Fusion Protein

This material was generated by Dr. Denish Piya [137]. The method is as follows: Synthetic DNA fragments that encompassed regions of plasmid pBAD24 from 50 bp upstream of the NheI restriction site to 50 bp downstream of the HindIII restriction site, with some modifications, were used. The ribosome binding site was modified to "aggaggt", followed by eight arbitrary nucleotides. The nucleotides encoding the N-terminal 9 or 30 residues of DarB (including the first residue Met) were added downstream of the arbitrary nucleotides. The pBAD24 multiple cloning sites from EcoRI to HindIII were added downstream of the *darB* specific nucleotides. Both synthetic DNA fragments were PCR amplified and ligated into NheI and HindIII restriction sites of pBAD24 using standard molecular biology techniques. TetR-mCherry [167] was PCR amplified with primers designed to have EcoRI and KpnI restriction sites on the 5′ end and ligated into pBAD24 to generate pBAD24_N30-tetR-mCherry, pBAD24_N9-tetR-mCherry, and pBAD_tetR-mCherry.

Generation of N-terminal DarB Mutants

This material was generated by Dr. Denish Piya ([137]. For constructing plasmid expressing *darB* with N-terminal truncations, the start codon ATG was included in the forward primers. The 5′ end of forward and reverse primers were designed to include XbaI and HindIII restriction sites. The amplified PCR product was ligated into XbaI and HindIII sites in pBAD24 following the manufactures protocol.

Generation of Internal DarB Truncation Mutants

Internal truncation mutants of *darB* were developed using Q5® Site-Directed Mutagenesis Kit (New England Biolabs). Primers were designed to truncate the N-terminal sequence and the amplicon was cloned into pBAD24 using the XbaI and HindII sites following the manufactures suggestions. The internal truncations were generated using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) using pBAD24_darB as a template. The amplicons were gel purified and used to transform 5-alpha competent *E. coli* (NEB) by heat shock and plated on selective media.

Site-Directed Mutagenesis of DarB

Internal truncation mutants of DarB were developed using Q5® Site-Directed Mutagenesis Kit (New England Biolabs). Primers were designed to make a point mutation in the codon of interest. The internal truncations were generated using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) using pBAD24_*darB* as a template following the manufacture's protocol. The amplicons were gel purified and used to transform 5-alpha competent *E. coli* (NEB) by heat shock and transformants were plated on selective media. Single colonies were picked and patched onto selective media.

SDS-Page Analysis

For SDS-PAGE analysis, samples were prepared as described previously [129], with slight modification [163]. Approximately 2×10^{10} plaque-forming units of CsCl purified P1 were loaded per lane and normalized to band intensity of the tail sheath protein. SDS-PAGE was conducted using 4–20% Bis-Tris SDS-PAGE gels (Thermo Fisher Scientific) with 5 uL PageRuler Unstained Broad Range Protein Ladder (Thermo

Fisher Scientific) and stained with SYPRO Ruby (Thermo Fisher Scientific) following the manufacturer's recommended protocol for maximum sensitivity. The gels were imaged with a Fotodyne gel imager.

Purification of Virions by CsCl Isopycnic Centrifugation

Purification of virions by CsCl isopycnic centrifugation was performed as previously described [129]. P1 or P1 mutants were induced in 1 L LB cultures as described above. 1L lysate was concentrated by centrifugation at 10,000 x g for 24 hours. The supernatant was discarded, the pellet was hydrolyzed with SM buffer (0.1 M NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5), and allowed to incubate at 4°C for 48 hours. The pellet was pipette mixed to homogeneity and 0.75 g/mL CsCl was added to the phage suspension. The phage suspension was transferred to Quick-Seal Ultra Centrifugation Tubes (Beckman Coulter) and sealed. Isopycnic gradient centrifugation was conducted in a Beckman 70.1 Ti rotor at 42,000 rpm for 24 h at 4 °C [163]. Phage bands were extracted with 18-gauge needles and dialyzed against 1 M NaCl SM buffer (1 M NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5) for 24 h in Slide-A-Lyzer 3500 MWCO dialysis cassettes (Thermo Fisher Scientific), then dialyzed against fresh SM buffer for another 24 h. Phage suspension was collected from the Slide-A-Lyzer and stored at 4 °C for further use.

P1ΔdarB Transformation and pBAD24_N30-tetR-mCherry Induction

A lysogen of P1Δ*darB* [129] was grown in LB Cm-R Kan-R to an OD550 of between 0.3 and 0.5. One mL of the liquid culture was centrifuged at 13,000 rpm in a benchtop microcentrifuge (Eppendorf Centrifuge 5424) for 1 min and the supernatant was discarded. The bacterial pellet was resuspended and washed in 1 mL ice-cold $_{dd}$H2O

twice, followed by one wash in ice-cold 10% (v/v) glycerol in $_{dd}H2O$ and resuspension in 100 µL of ice-cold 10% glycerol in $_{dd}H2O$ containing 1 µL of either pBAD24_darBN30 or pBAD24_darBN9. The solution was transferred to an ice-cold 1 mm electroporation cuvette and electroporated using a Bio-Rad Micro-PulserTM following the manufacturer's protocol. After electroporation, 900 µL of LB was added to the cuvette and the suspension was incubated at 30 °C for 1 h. The culture was centrifuged at 13,000 rpm, the supernatant was discarded, and the bacterial pellet was resuspended in 200 µL of LB. Then, 100 µL of the bacterial suspension was spread onto an LB Kan-R Amp-R agar plate, allowed to dry, and placed in a 30 °C incubator overnight. Transformant colonies were inoculated into LB Kan-R Amp-R, incubated overnight, subcultured 1:100 into 1 L LB Kan-R Amp-R and allowed to grow to an OD_{550} of 0.3–0.5. The cultures were then induced with 1 mM L-arabinose (Sigma-Aldrich), shifted to 42 °C, and incubated until the OD_{550} was below 0.2. The remaining cells were lysed by the addition of 0.1% (v/v) chloroform and the supernatants purified by isopycnic CsCl gradient centrifugation as described above.

Efficiency of Plating (EOP) Assays

All phages used in EOP assays were lysogenized into *E. coli* WA921 and phages produced by induction as previously described [129]. EOP assays were conducted following previously established protocol, with minor modifications [129]. P1 Phages were serially diluted 10-fold in SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5). 4 mL of molten T-Top agar (LB with 0.5% agar) with a final concentration of $CaCl_2$ of 5mM and 100 uL of bacteria at OD_{550} 0.3-0.5 was poured onto LB agar

plates and allowed to solidify. 10 μL from each dilution was spotted on the soft agar overlays and allowed to dry. Experiments were conducted in triplicate.

Fluorescence Microscopy Imaging

Imaging was performed on a Nikon Eclipse Ti inverted epifluorescence microscope using a 100x objective (Plan Fluo, NA 1.40, oil immersion) with a 2.5x TV relay lens, within an incubator cage (InVivo Scientific) at 30 °C and images acquired using a cooled EMCCD (electron multiplying charge-couple device) camera (IXON 897, Andor, Belfast, UK). DAPI and mCherry images were collected using filter cubes for DAPI (Nikon no. 96310) and ET/mCH/TR (mCherry/Texas Red, Nikon no. 96365), respectively. Images were collected from 16 random fields and correlation of DAPI (400 ms exposure) and mCherry (500 ms exposure) foci was conducted manually in the NIS-Elements software (Nikon).

Results

BLAST of DarB Revels Helicase and Methyltransferase Domains

A key enzyme in the anti-restriction system of P1 is DarB as it confers protection against *E. coli* type I endonucleases EcoB and EcoK [129]. While the mechanism of protection is not well understood, protein BLAST is a valuable tool for identifying motifs in an amino acid sequence. Protein BLAST revealed DarB to have helicase and methyltrasnferase domains which suggest a mechanism of action (Figure 11).

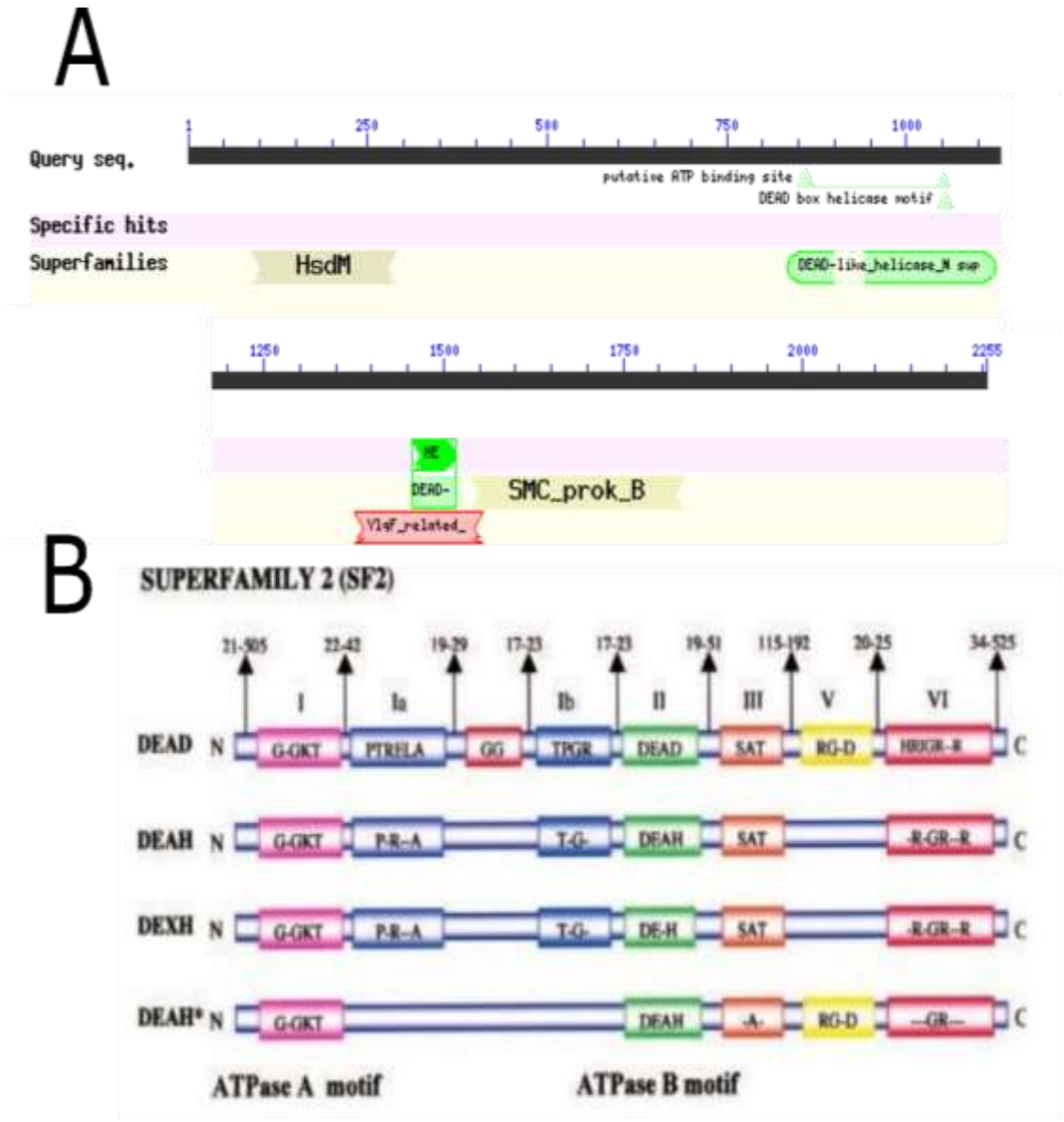

Figure 11 Blastp output of DarB primary structure and Superfamily 2 motifs. A) The primary sequence of DarB was used as in input for protein-protein BLAST in NCBI. The important superfamilies identified are HsdM and Dead-like helicase motifs. Additionally, BLAST revealed DarB to harbor ATP-binding sites within the dead-like helicase domain. B) Dead-like helicase motifs form variants of the dead-like domain. DarB is part of the DEXH subfamily.

Site-Directed Mutagenesis Identifies Key Residues Important For Anti-Restriction Activity

The methyltransferase domain of DarB has homology to the methyltransferase domain of type I *E. coli* restriction-modification system. Additionally, the helicase domain of DarB is part of helicase Superfamily 2 [168]. Specifically, DarB has motifs from the DEXH variant. The motifs from superfamily two are well defined and mutations within those motifs do not affect the tertiary structure of the helicase. [168]. Site-directed mutagenesis was used to understand the importance of the catalytic arginine within the helicase motif of DarB (Figure 12).

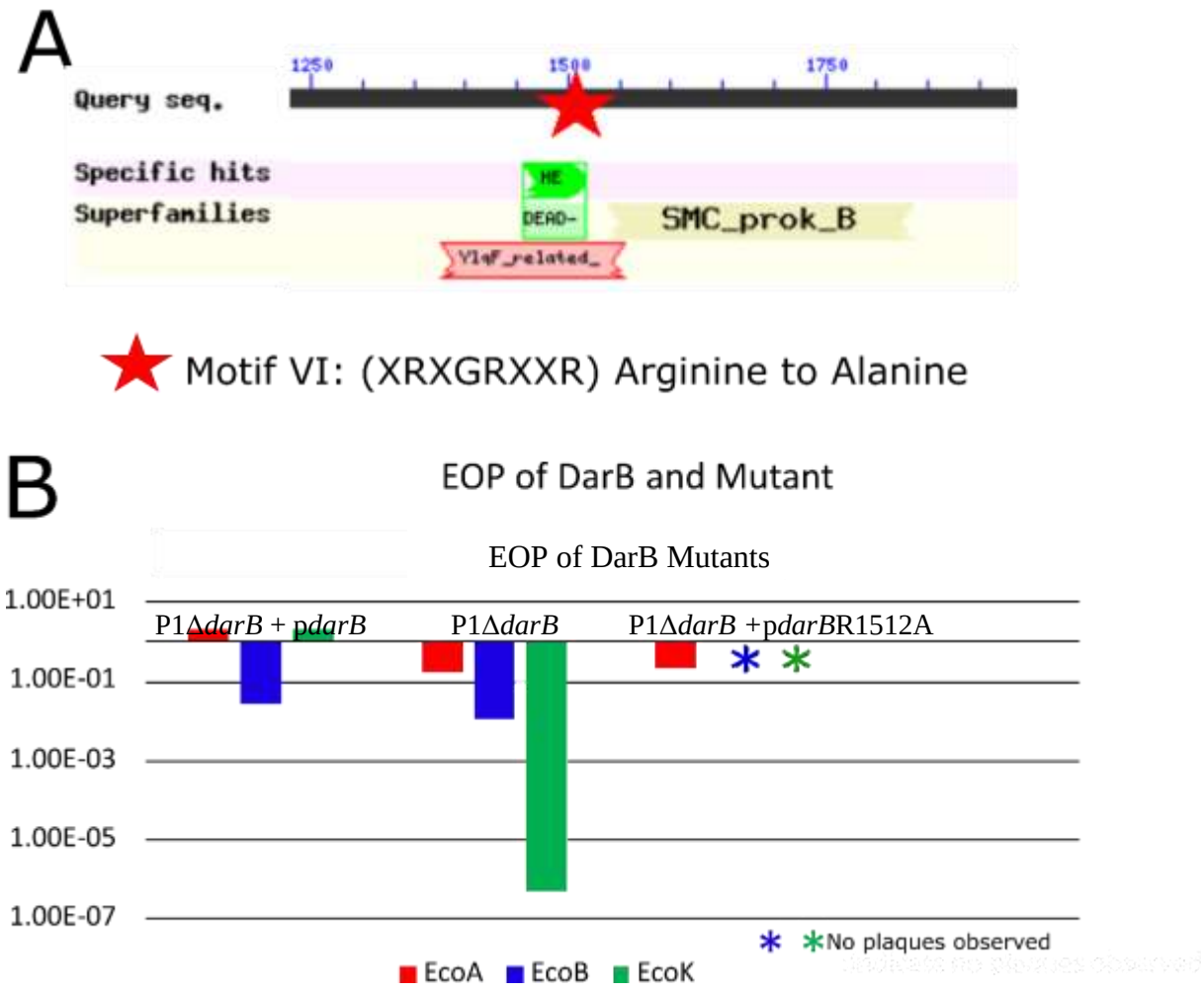

Figure 12 BlastP output of DarB primary sequence and efficiency of plating of DarB point mutant. A) The primary sequence of DarB was used an in input for protein-protein BLAST in NCBI. The red star on the query sequence indicates the location of the point mutation. The second arginine in the primary amino acid sequence provided is mutated in a non-synonymous fashion to encode for the amino acid alanine. B) The bar graph indicated the changes in efficiency of plating between P1**Δ***darB* + p*darB*, P1**Δ***darB*, and P1**Δ***darB* + p*darB*R1512A. The colors of the bars in the graph are indicative of a specific restriction enzymes identified in the legend. Additionally, no plaques were identified in the type VI point mutant which is indicated by asterisks (*) replaces the bars.

The point mutation in motif VI maintained DarB packaging while also abolishing the anti-restriction activity of DarB. Further, site-directed mutagenesis was used to mutate methyltransferase motifs. One methyltransferase mutant revealed an incomplete inhibition of DarB anti-restriction activity showing EcoB protection to be that of a DarB knockout, whereas EcoK protection was maintained (Figure13). While packaging of these DarB mutants was not confirmed, the mutation in the methyltransferase motif has nearly wild-type protection from restriction. This suggests that the mutant DarB is packaged.

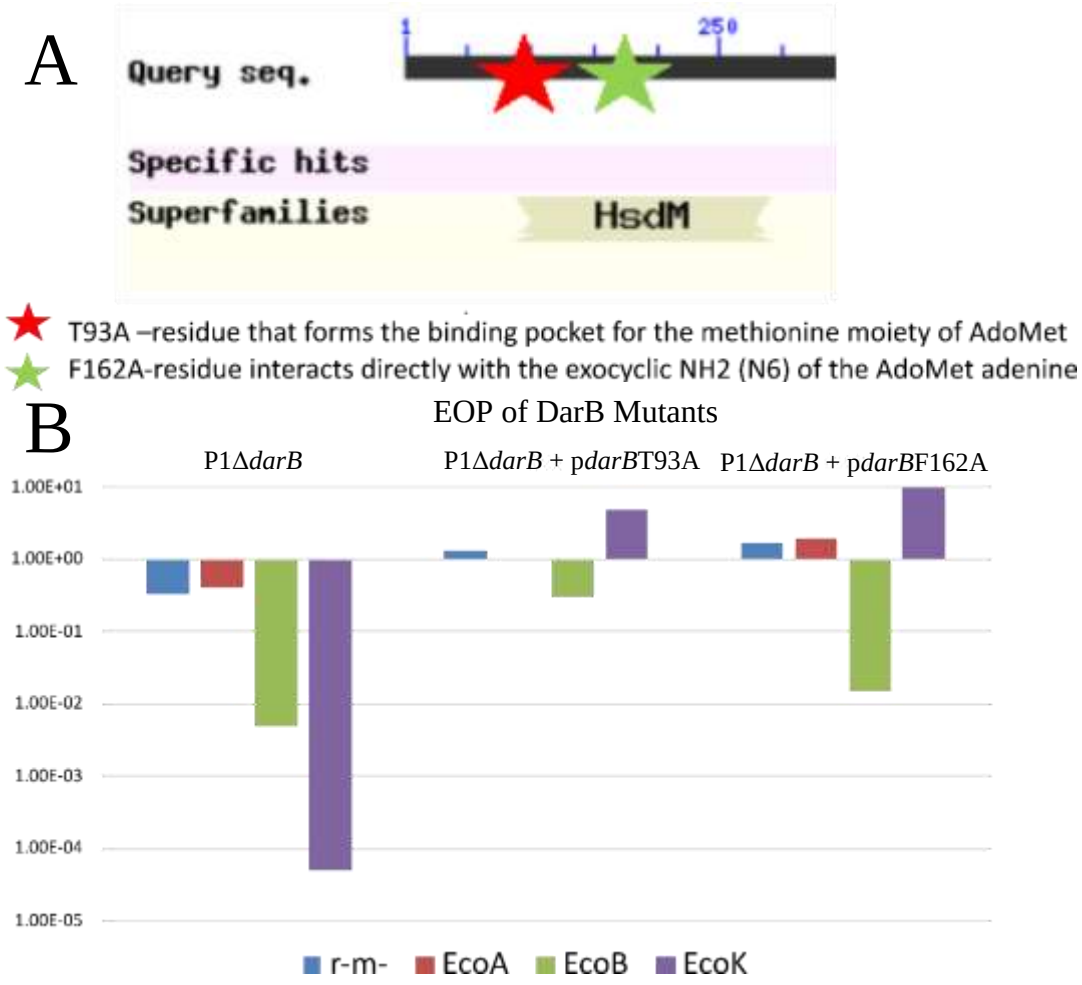

Figure 13 BlastP output of DarB primary sequence and efficiency of plating of DarB point mutant. A) The primary sequence of DarB was used an in input for protein-protein BLAST in NCBI. The truncated BLAST output is shown. The red star on the query sequence indicates the location of the T93A point mutation. The green star on the query sequence indicates the location of the F162 point mutation. B) The bar graph indicated the changes in efficiency of plating between P1**Δ***darB*, P1**Δ***darB* + p*darB*T93A, and P1**Δ***darB* + p*darB*F162A. The colors of the bars are indicative of a specific restriction enzymes identified in the legend.

DarB Truncation Mutants Reveal Capsid Localization Signal of DarB

P1 encodes the multi-component defense against restriction (Dar) system that protects the P1 chromosome from host Type I restriction upon infection [94]. This system consists of six known proteins that are incorporated into the P1 virion in a stepwise process, in which the 250 kDa DarB protein is the last Dar component added to the capsid [129]. Phage T4 is understood to package internal proteins into its capsid, which are delivered to the host cell upon infection and play a role in defeating host restriction endonucleases [169]. Mullaney and Black [36] described a capsid targeting sequence (CTS) in the internal head protein IPIII of phage T4 that is sufficient to target foreign proteins to the T4 capsid while retaining their function. To determine if DarB is directed to the P1 capsid by a similar mechanism, truncation mutants at the DarB N-terminus were constructed.

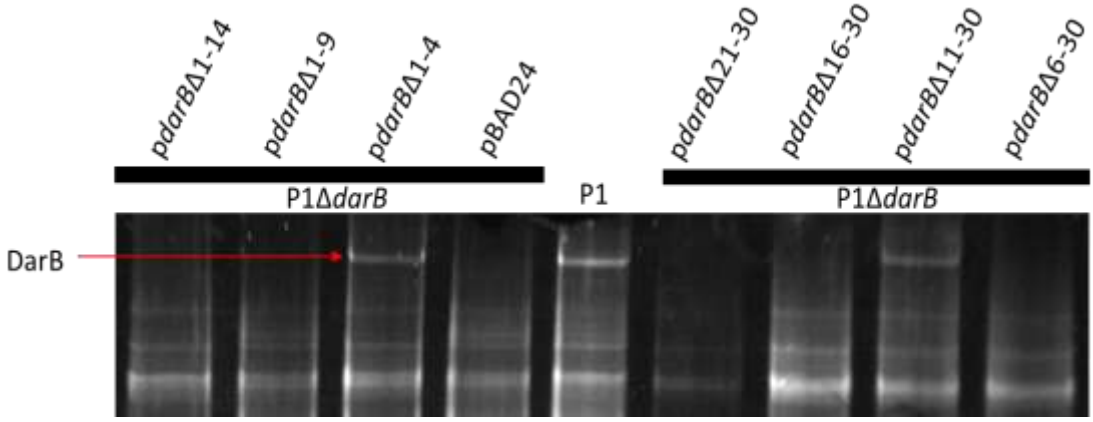

Figure 14 SDS-PAGE identifies DarB capsid localization signal. N-terminal truncations were cloned into pBAD24 and expressed in trans. The band shown is SYPRO-RUBY strained DarB. DarB is only observed in the wild-type, 1-4, and 11-30 truncation backgrounds.

Capsid Localization Signal of DarB Packaged mCherry to the P1 Virion

Figure 14 shows the ability of DarB truncations to localize to the P1 capsid. Truncations of DarB residues 1-4 and 11-30 are tolerated, which would suggest amino acids 6 through 10 are required to localize DarB to the P1 virion. However, deletion of residues 16-30 and 21-30 also inhibit DarB packaging, suggesting that other deletions may induce improper folding of the N-terminus that inhibits recruitment of DarB to the P1 virion (Figure 14).

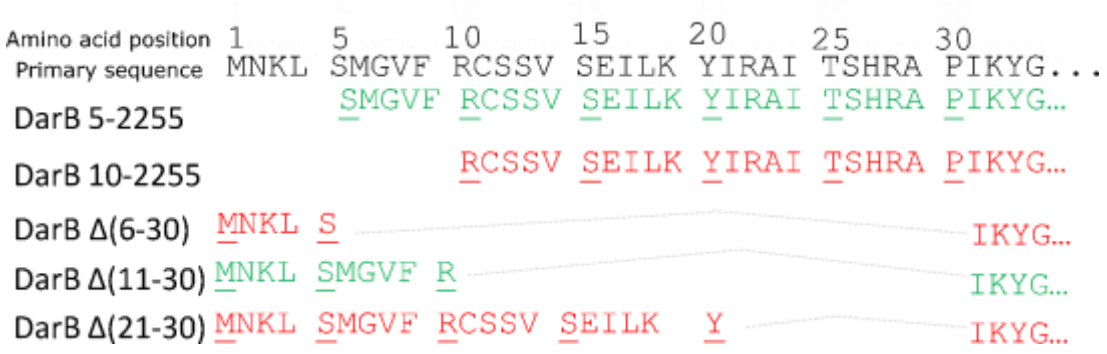

Figure 15 Alignment of the amino acid primary sequence of select DarB truncations. The primary sequences colored in green indicate the truncations that are packaged within the P1 virion, while the primary sequences in red indicate the truncations packaged within the P1 virion. The hashed lines indicate the missing amino acids. The underlined letters of within the primary sequences of the truncated backgrounds is meant to serve as a marker of every 5th amino acid.that are not packaged within the P1 virion. The hashed lines indicate the missing amino acids. The underlined letters of within the primary sequences of the truncated backgrounds is meant to serve as a marker of every 5th amino acid.

To further examine the DarB CTS, we wished to determine if this signal is sufficient to target a heterologous protein to the P1 capsid. To this end, the N-terminal 30 residues of DarB were fused to the N-terminus of the fluorescent reporter TetR-

mCherry [167] and cloned into pBAD24 to generate the expression vector pBAD24_N30-tetR-mCherry. Next, a P1Δ*darB* lysogen was transformed with pBAD24_N30-tetR-mCherry. The P1Δ*darB* prophage and plasmid were simultaneously induced to produce P1Δ*darB*+TetR-mCherry virions. Finally, to visualize the localization of the TetR-mCherry to the P1 virion, virions were purified in CsCl isopycnic gradients and stained with the DNA intercalating dye 4',6-diamidino-2-phenylindole (DAPI). As shown in Figure 16A, the DAPI and mCherry foci of cesium chloride isopycnic centrifugation purified P1ΔdarB+TetR-mCherry virions are easily distinguishable in fluorescence microscopy. Fluorescence microscopy images were taken over a 10-day period and ~500 foci were counted in each DAPI and mCherry channel on each of days 0, 4, and 10. While DAPI foci and mCherry foci colocalize above 90% over the ten-day period (Figure 16B, blue line), mCherry foci colocalize with DAPI foci at nearly 60% on day 0 which declines to approximately 30% on day 10 (Figure 16B, red line). P1ΔdarB was also induced in the presence of TetR-mCherry lacking the DarB CTS, and this construct produced P1 virions lacking fluorescent signal (Figure 16), indicating the CTS is required for TetR-mCherry packaging in the P1 virion. This process was also conducted with the first 9 residues of DarB fused to the N-terminus of TetR-mCherry, but the fusion construct did not target fluorescent signal to the P1 virion, indicating the complete CTS is greater than 9 residues and less than 30 residues in length.

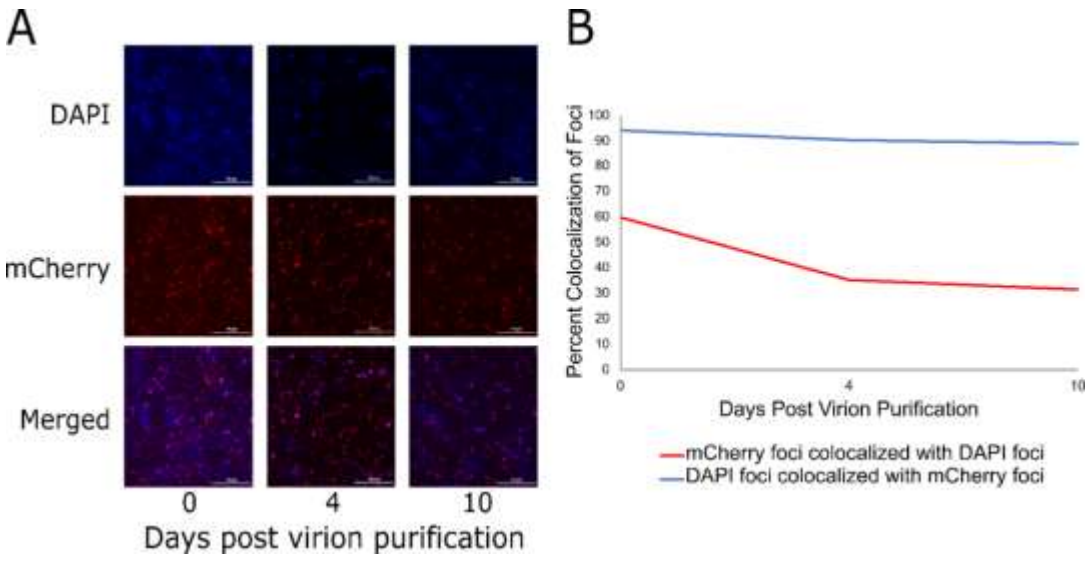

Figure 16 The 30 N-terminal residues of DarB are sufficient to target TetR-mCherry to the P1 capsid. A) The simultaneous induction of the pBAD24_N30-tetR-mCherry expression vector and the P1Δ*darB* lysogen results in the localization of TetR-mCherry to the P1 capsid. Fluorescence imaging of CsCl-purified P1Δ*darB*+ TetR-mCherry purified virions over days 0, 4, and 10 are shown using a DAPI filter, mCherry filter, and the two channels merged. B) The line graph illustrates the colocalization of DAPI and mCherry foci over the 10 days post-purification of the P1Δ*darB*+ TetR-mCherry virions. The red line indicates the percent colocalization of mCherry foci and DAPI foci, and the blue line indicates the percent colocalization of DAPI foci with mCherry foci. This indicates that most (~90%) particles with packaged DNA contain TetR-mCherry, but many particles containing TetR-mCherry do not contain DNA; this may be due to particle instability or premature DNA ejection from the virions during storage.

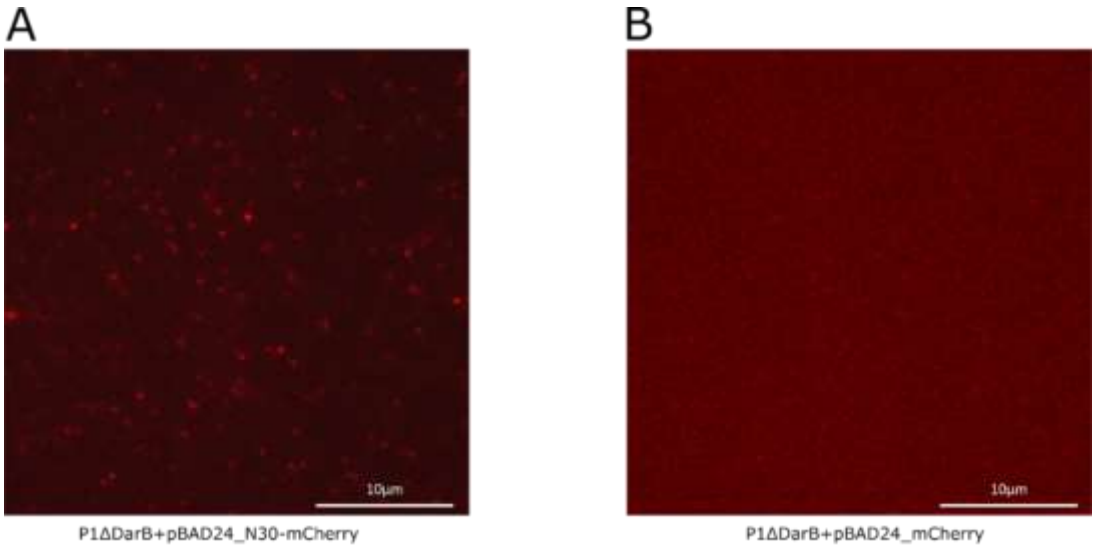

Figure 17 P1Δ*darB*+pBAD24_N30-mCherry and P1Δ*darB*+pBAD24_mCherry virions were purified by side-by-side by CsCl isopycnic centrifugation and applied to slides for observation by fluorescence microscopy. A) Fluorescence imaging of P1Δ*darB*+pBAD24_N30-mCherry shows fluorescent foci using the mCherry filter. B) Fluorescence imaging of P1Δ*darB*+pBAD24_mCherry shows the absence of fluorescent foci using the same filter. Camera signal gain was set automatically for image capture; therefore, the image background is amplified in (B), which appears dark under manual observation.

The exogenous packaging of TetR-mCherry to the P1 virion is similar in principle to the protein expression, packaging, and processing (PEPP) system of phage T4 developed by Mullaney et al. [170]. The T4 PEPP system has been shown to target diverse substrates to the T4 capsid via its N-terminal packaging signal, including HIV-1 protease, micrococcal endonuclease from Staphylococcus aureus, restriction endonuclease EcoRI, luciferase, human granulocyte colony stimulating factor (GCSF), green fluorescent protein (GFP), and the 99 amino acid C-terminus of amyloid precursor

protein (APP) [170]. The P1 capsid targeting sequence described here appears to serve an analogous function as the capsid localization signal described for the T4 PEPP system. The P1 system, however, has the advantage of being coupled to a much broader host range phage than T4; P1 has been shown to infect hosts as distantly related as Myxococcus [171] suggesting the P1 system could be employed to deliver proteins to a wide variety of bacterial targets.

I quantified the colocalization of DAPI foci and mCherry foci to be 90%, while the number of mCherry foci and DAPI foci are initially 60% on day zero and this declines to 30% at day 10. The drop in colocalization of mCherry foci and DAPI foci over the ten-day period is likely due to auto-ejection of the DNA during storage. The steady colocalization of DAPI foci and mCherry foci over the ten-day period is understandable as these are assumed to be virions that do not have their ejection mechanism disturbed during storage, thus the capsid harbors DNA and TetR-mCherry. It is reasonable to not expect 100% colocalization of DAPI foci and mCherry foci since there is a minimum amount of TetR-mCherry that needs to be packaged in the P1 virion to observe fluorescence. Previous work indicated that ~40 copies of DarB are packaged per virion in bulk measurements [129], but this number may vary widely between individual virions. Another possibility is that the TetR-mCherry protein encounters difficulty in folding while packaged into the P1 virion. Additionally, a band containing only P1Δ*darB*+TetR-mCherry tail-less heads was collected from the CsCl gradient and imaged by fluorescence microscopy on day 0 and day 10 (Figure 18A). The purified P1Δ*darB*+TetR-mCherry heads display a 90% colocalization over the 10-day period

(Figure 18B) which suggests that the low rate of colocalization in the purified virions is due to an instability in the P1 DNA ejection mechanism mediated by the presence of the functional phage tail.

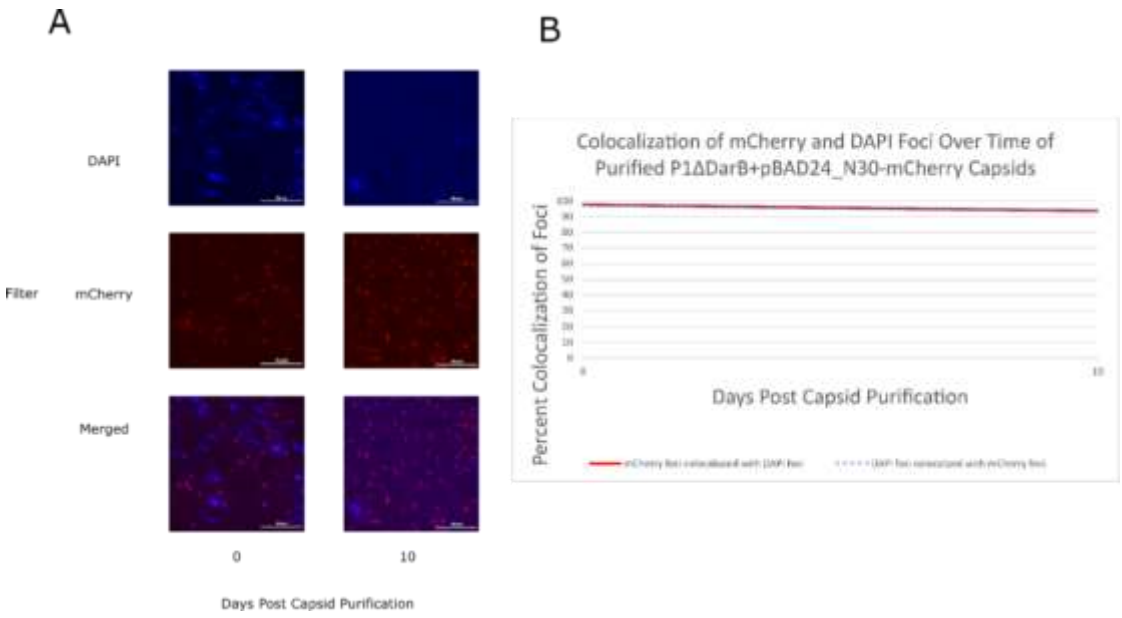

Figure 18 P1ΔdarB capsids purified by cesium chloride isopycnic centrifugation. A) The simultaneous induction of the pBAD24_N30-mCherry expression vector and the P1Δ*darB* lysogen results in the localization of mCherry to the P1 capsid. Fluorescence imaging of cesium chloride isopycnic centrifugation P1Δ*darB*+pBAD24_N30-mCherry purified heads on days 0 and 10 are shown using a DAPI filter, mCherry filter, and the two filters merged. B) The line graph illustrates the colocalization of DAPI and mCherry foci over the 10 days post cesium chloride isopycnic centrifugation of the P1Δ*darB*+pBAD24_N30-mCherry heads. The red line indicates the percent colocalization of mCherry foci and DAPI foci. Here, colocalization of foci occurs when an mCherry foci is observed first, then, secondly, the filter is switched to DAPI and the presence of a DAPI foci overlaps with the initial mCherry foci. The blue dotted line indicates the percent colocalization of DAPI foci and mCherry foci. Here, colocalization of foci occurs when a DAPI foci is observed first, then, secondly, the filter is switched to mCherry and the presence of an mCherry foci overlaps with the initial DAPI foci.

It is reasonable to suspect that Dar proteins share a common CTS, but BLAST analysis does not reveal a consensus sequence among the N-termini of Dar proteins.

While the data from individual truncations suggests the CTS for DarB lies in residues 6-10, the first 30 amino acids are shown to be sufficient to direct the reporter TetR-mCherry to the P1 capsid. It is possible that the N-terminus of DarB requires proper tertiary structure to localize to the P1 capsid, as certain N-terminal truncations could compromise folding at the N-terminus and inhibit capsid targeting.

Conclusion

Bacteriophages must overcome their host's defense systems in order to be infectious. The multicomponent restriction system harbored by P1 not only protects the packaged P1 DNA from *E. coli* type I restriction endonucleases, but also protects the P1 DNA from *Salmonella* type I restriction endonucleases [94]. Further, the anti-restriction phenotype of P1 is not limited to P1 DNA, but rather the packaged DNA within the P1 capsid during infection. Unlike the anti-restriction activity of T7's Ocr which mimics DNA, the anti-restriction mechanism of the P1 Dar system is not understood. A key protein shown to exhibit protection from EcoK and EcoB is P1's anti-restriction protein DarB. It has been previously shown that the Dar system is incorporated into the P1 virion in a stepwise manner with DarB being added to the P1 virion last making DarB a prime candidate for further genetic studies as mutations would not affect packaging of the defense against restriction system into the P1 virion. Using the NCBI BLAST tool, DarB was shown to have homology to type III endonuclease subunit M. The M subunit of type III endonucleases is responsible for modifying the bacterial chromosome after DNA replication serving as a methyltransferase. Further, DarB has a DEXH helicase motif and ATP binding sites suggesting DarB is able to exhibit helicase activity in an

ATP-dependent manner. Taken together, DarB is speculated to modify the P1 DNA, which protects it from bacterial type III endonucleases. It was hypothesized that point mutations within the active sites of motifs within the methyltransferase and helicase domains would decrease the anti-restriction activity of DarB. To this end, site-directed mutagenesis was employed on key motifs in the methyltransferase and helicase domains. Mutant DarB cloned into pBAD24 was used to transform P1 *darB* knockout lysogens in order to make an *in trans* expression system (WA921 + P1Δ*darB*::Kan (pBAD24_*darB*X#Z) [X#Y is used to indicate the point mutation]) that would package the mutant DarB in the P1 virion. Efficiency of plating was used to compare P1Δ*darB*::Kan(pBAD24_*darB*R1512A) to P1Δ*darB*::Kan and P1Δ*darB*::Kan(pBAD24_*darB*) backgrounds. The expression vector with pBAD24_*darB*R1512A showed no plaques on EcoB or EcoK hosts, while showing near wild-type protection for EcoA (Figure 6). Motif VI is responsible for stabilizing ATP. Specifically, R1512 is responsible for stabilizing the gamma-phosphate in an effort to coordinate it with the magnesium cation to enable hydrolysis, thus enabling helicase activity in an ATP-dependent manner [168]. The complete ablation of the anti-restriction activity of DarBR1512A suggests that the anti-restriction activity of DarB requires helicase activity. While it is unknown if DarB exhibits DNA modification activity in the capsid or in the bacterial cytoplasm it is understood that DarB must have helicase activity to protect the packaged DNA from EcoB and EcoK restriction endonuclease activity. In T4 it has been shown that staphococcal nuclease is able to act on DNA packaged within the T4 capsid [36].

To explore the significance of the methyltransferase activity on the anti-restriction phenotype, DarB was mutated at amino acids T93 and F162. T92 forms the surrounding of the binding pocket for the methionine moiety of AdoMet which is transferred from AdoMet to the DNA, while F162 interacts directly with the nucleoside of AdoMet in an effort to stabilize the molecule. While there was no significant loss of anti-restriction activity from the T93 point mutation, the F162 point mutant exhibited decrease protection against EcoB while having greater than wild-type protection for EcoK. Increased protection against EcoB and EcoK has been previously reported for vectors expressing DarB *in trans*. This is suspected to be due to increased packaging of DarB in the P1 virion and thus a greater capacity of protection [129]. This previous observation could explain why DarBT162A has greater protection against EcoK. However, the lack of protection from EcoB from DarBT162A is interesting. DarBT162A still has helicase activity, but the methyltransferase activity is not completely abolished since there is still protection against EcoK. Type I restriction systems rely on a recognition subunit which recognizes unmethylated DNA. When unmethylated DNA is recognized, the type I restriction system binds to the DNA, traverses the DNA and exhibits hydrolysis of the DNA substrate. The presence of EcoK protection suggests that EcoK is unable to bind to P1 DNA while the lack of protection from EcoB suggests that EcoB is able to bind to the P1 DNA. This “chimeric” protection of the P1 genome from DarBT162A suggests that only a fraction of the P1 genome is modified. Indeed, within the genome the number of recognition sequences for EcoB is 20 while the number of recognition sequences for EcoK is 11. There are nearly 50% less recognition sequences

for EcoK than there are for EcoB. The data suggests, that DarBT162A is able to methylate the P1 DNA, but not to such an extent to inhibit binding of the EcoB to the P1 DNA during infection.

Next, experiments were conducted to further understand the capsid localization signal of DarB. T4's IPIII has been shown to harbor a capsid localization signal. Further, the capsid localization was used to target MNase to the T4 capsid with success. Similarly, I wanted to identify the capsid targeting sequence of DarB to near amino acid precision. To better understand the amino acids responsible for packaging DarB to the P1 virions, a series of truncation mutants as well as internal truncations were made (Figure 14). Vectors expressing truncated DarB were used to transform P1Δ*darB*::Kan to make P1Δ*darB*::Kan(p*darB*Δ# ["#" indicates the amino acids absent from the DarB protein]). Transformed P1Δ*darB*::Kan were induced for virion assembly and vector expression simultaneously and the virions were purified by isopycnic ultra-centrifugation. Purified virions of truncation mutants were observed by SDS-PAGE to visualize the packaging of DarB (Figure 14). In this way, it was observed that all truncation mutants except *darB*Δ11-30 and *darB*Δ1-4 localized DarB to the P1 virion. This suggests that amino acids 5-10 are the capsid localization signal of DarB. The capsid localization signal of DarB is not shared with other anti-restriction proteins. This is understandable as the stepwise incorporation of the defense against restriction system of P1 requires precursors to be packaged before proteins can be target to the P1 capsid suggesting there may be unique quaternary features used by each component of the anti-restriction system. Further, the suggested capsid localization signal DarB presented by

this works spans a length of appears to be 6 amino acids. This is in contrast with the capsid localization signal of IPIII of T4 which has a capsid localization signal comprised of the first 10 amino acids of IPIII. Since the capsid localization signal of IPIII was shown to package exogenous proteins to the T4 virion, the capsid localization signal of DarB was tested to observe if it would package a foreign protein to the P1 capsid.

The capsid localization signal of T4's IPIII has been shown to package staphylococcal nuclease, *Eco*RI endonuclease, β-globin, and luciferase. In the case of P1, the capsid localization signal was fused to mCherry. To explore if the capsid localization signal of DarB is able to package mCherry, P1Δ*darB*::Kan lysogens were transformed with pN30-tetR-mCherry and the lysogen and expression vector were simultaneously induced. Next, virions were purified by cesium chloride isopycnic ultra-centrifugation and the purified virions were observed by fluorescent microscopy. Fluorescence microscopy demonstrated that mCherry fused to the first 30 amino acids of DarB was localized to the P1 capsid. To affirm if the observed mCherry foci were, in fact, located in the P1 virion, purified virions were stained with DAPI. A positive identification of DarB within the P1 virion was considered to be observed when an mCherry foci coincided with a DAPI foci. Curiously, over the course of 10 days, DAPI foci colocalized with mCherry at nearly 100% coincidence. The lack of a 100% coincidence of DAPI to mChery foci may be explained by an insufficient amount of mCherry is packaged in nearly 1-out-of-20 virions. On the other hand, over the 10-day period mCherry foci colocalized with DAPI started at 60 percent and decreased to 40 percent. The decrease in mCherry to DAPI colocalization over the 10-day period is

suspected to be due to auto-ejection of DNA. This is supported by the data which shows that purified P1 capsids were observed to have 100% colocalization of mCherry to DAPI and DAPI to mCherry (Figure 18). Since the capsids are missing a tail, there appears to be no trigger present to eject the contents of the capsid. If the DNA is auto ejected and mCherry is still located in the P1 capsid, this suggests that mCherry cannot be ejected from the P1 virion during this process. Further, if mCherry cannot be ejected from the P1 virion with a mass of 26.7 kDa, then DarB with a mass of 250 kDa will not be ejected from the P1 capsid either. This may suggest that the activity of DarB occurs within the P1 capsid, however this observation may be explainable by other methods, such as incomplete ejection of protein from the capsid with DNA, or differences in the process of auto-ejection compared to ejection during cell infection.

The data presented here demonstrates the requirement of DarB helicase and methyltransferase domains to be important for anti-restriction activity. Specifically, the helicase domain is imperative to the anti-restriction activity of DarB. Further, mutations in the methyltransferase domain were shown to decrease the anti-restriction activity of EcoK, but not EcoB. The discrepancy between anti-restriction activity is presumed to be due to the number of recognized sequences for EcoB and EcoK within the P1 genome. Further study involving changing the amount of EcoB sites present in the P1 genome will shed light on the notion that protection against EcoK is due to the amount of EcoK is due to the low abundance of recognition sequences found in the P1 genome. Next, the identification of the capsid localization sequence of DarB was suggested to be amino acids 5-10, but amino acids 1-30 were capable of packaging mCherry to the P1 capsid.

This discrepancy of the capsid localization signal may be due to other regions of DarB interacting with the quaternary structure of the P1 capsid to enable packaging. Still, mCherry was able to be packaged into the P1 virion using the first 30 amino acids of DarB which provides a precedence to package other exogenous proteins to the P1 capsid. Further, it was observed that DAPI to mCherry foci colocalization was near 100% after 10-day storage at 4°C, but mCherry to DAPI foci colocalization was nearly 40% after 10 days of storage at 4°C. This may be due to auto ejection of phage DNA and further suggests that mCherry can not be expelled from the P1 virion, or is expelled inefficiently during this process. As suggested earlier, if mCherry can not be expelled from the P1 capsid, then DarB would not be able to be expelled from the P1 capsid either. If this is the case, then DarB would have to exhibit its activity within the P1 capsid. This would be easily tested by contrasting sequencing data between wild type P1 virions and P1 virions lacking DarB which were both grown in modifications deficient hosts. Ideally, the modification of DNA packaged within the P1 capsid would have to occur after packaging of DNA to the P1 capsid with the capsid still located within the host cytoplasm. Presumably the amount of ATP and AdoMet are able to enter the capsid through osmosis and be consumed by the anti-restriction components of P1. While staphylococcal endonuclease was able to act on packaged T4 DNA within the T4 capsid, this enzyme does not require cofactors to function, which would be a limitation for DarB based on its predicted requirement for both ATP and a methyl group donor.

CHAPTER IV

CRYO-EM OF BACTERIOPHAGE P1 REVEALS A NOVEL CENTRAL SPIKE PRONG ENCODED BY UPFB WHICH PLAYS A ROLE IN INFECTION

Introduction

Bacteriophages are the most abundant life on earth, with a majority of phages belonging to the morphotype myophage [29]. Myophages are phages that have a capsid attached to a rigid contractile tail. The molecular structure of the contractile tail is well understood for phage T4, a myophage model system [162]. Unlike other any phage, including P1, phage T4 has been studied extensively by x-ray crystallography and cryo-electron microscopy.

Many of the proteins identified in T4 have homologs in other myophage [162]. Specifically, components of the T4 baseplate have been observed in other myophages as well as the *E. coli* type VI secretion system. Due to the conserved nature of the T4 baseplate it has been called a "conserved core" [162]. The T4 conserved core contains proteins that are also found in R-type pyocins and the bacterial type 6 secretion system [162]. A simpler baseplate core has been identified in phage Mu and homologs have been identified in T4 and the T6SS [50]. In contrast, while phage P1 has been used as a genetic workhorse for mapping the *E. coli* genome, structural work on P1 slowed in the mid 1980's. With the advent of molecular cloning, and centers established to aid laboratories in high-resolution microscopy, phage P1 should be further examined in a similar way that T4 has. Phage P1 is of interest due to its transducing capacity for *E. coli* as well as its capability to infect various other *Enterobacteriaceae* [166]. The broad host range of

P1 is widely appreciated and I hypothesize that characteristics of the P1 baseplate may contribute to the broad host range.

The functional annotation of phage P1 has a large number of genes without well-defined functions [96]. The majority of unknown genes have been found to be non-essential for assembly and infection [172]. P1 structural proteins homologous to the T4 common baseplate core have been identified [172]. Still, there is lacking a detailed structural exploration of P1 structure using modern methods. Here, cryo-EM has been used to reconstruct the baseplate of phage P1. Further, baseplate proteins have been modeled by AlphaFold to reconstruct the protein folds and mapped onto the density map revealed by cryo-EM. This work emphasizes a heretofore unrecognized P1 central spike with multiple prongs, proposes the gene encoding the prong, and produces the first density map of phage P1 baseplate and lower tail.

Materials and Methods

Bacterial Strains and Phages

The bacterial strains and the parental phage P1CMclr100 (hereafter referred to as P1) used in this study were obtained from the Coli Genetic Stock Center, Yale University, from our previous studies, or obtained from the Provenzano lab stock at The University of Rio Grande Valley - Brownsville Campus [172]. Unless otherwise noted, *E. coli* strains, *V. cholerae* strains, *Salmonella typhimurium* strains, *K. aerogenes*, *P. mirabilis*, and *S. marcescens* were cultured in LB (10 g/L Bacto Tryptone (BD Biosciences), 5 g/L Bacto yeast extract (BD Biosciences), 10 g/L NaCl (Avantor)) or LB agar (LB amended with 15 g/L Bacto agar) and incubated at 37 °C. P1 lysogens were

cultured and maintained at 30°C on LB amended with 10 ug/mL chloramphenicol (LB Cm-R) or 10 ug/mL chloramphenicol plus 30 ug/mL kanamycin (LB Cm-R Kan-R).

Production of Phage Lysates

Briefly, an *E. coli* P1 lysogen was grown at 30 °C in LB Cm-R to an OD550 between 0.5 and 0.6. Next, the P1 lysogen was thermally induced by shifting the culture to 42°C [94] until the OD550 fell below 0.2. Then, chloroform was added (0.1% v/v) to the lysate. Finally, the lysate was centrifuged at 10,000 x g for 30 minutes, and the supernatant was stored at 4 °C until further use.

Purification of Virions by CsCl Isopycnic Centrifugation

Purification of virions by CsCl isopycnic centrifugation was performed as previously described [129]. P1 or P1 mutants were induced in 1 L LB cultures as described above. 1L lysate was concentrated by centrifugation at 10,000 x g for 24 hours. The supernatant was discarded, the pellet was hydrolyzed with SM buffer (0.1 M NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5), and allowed to incubate at 4°C for 48 hours. The pellet was pipette mixed to homogeneity and 0.75 g/mL CsCl was added to the phage suspension. The phage suspension was transferred to Quick-Seal Ultra Centrifugation Tubes (Beckman Coulter) and sealed. Isopycnic gradient centrifugation was conducted in a Beckman 70.1 Ti rotor at 42,000 rpm for 24 h at 4 °C [163]. Phage bands were extracted with 18-gauge needles and dialyzed against 1 M NaCl SM buffer (1 M NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5) for 24 h in Slide-A-Lyzer 3500 MWCO dialysis cassettes (Thermo Fisher Scientific), then dialyzed against fresh SM

buffer for another 24 h. Phage suspension was collected from the Slide-A-Lyzer and stored at 4 °C for further use.

Lysogeny Establishment Assay

Phage lysate was used to infect *V. cholerae* strains and *S. marcescens* by combining 100 µL of bacteria at OD_{550} 0.3-0.5 with 100 µL of the phage lysate with a final concentration of 5 mM CaCl2. The phage and bacterial mixture were allowed to incubate at room temperature for 30 minutes. The infection mixture was centrifuged for 1 minute at 13,000*rpm, the supernatant discarded, and the bacterial pellet was resuspended in 100 µL LB. The resuspension was plated on LB agar plates containing 10 ug/mL chloramphenicol and incubated at 30°C overnight.

Phage Titer

Each P1 lysate background were serially diluted 10-fold using SM buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris-HCl pH 7.5). 4 mL of molten T-Top agar (LB with 0.5% agar) with a final concentration of $CaCl_2$ of 5mM and 100 uL of bacteria at OD_{550} 0.3-0.5 was poured onto LB agar plates and allowed to solidify. 10 µL from each dilution was spotted on the soft agar overlays and allowed to dry. Plates were allowed to incubate overnight at 42°C. Experiments were conducted in triplicate.

Cryo-EM Grid Preparation

QuantFoil 1.2/1.3 R200 grids were glow discharged as 15 mA for 60 seconds. Vitrobot was used with -15 blot force and 2.5 seconds of blot time. Grid freezing, and grid clipping was done at the Pacific Northwest Center for Cryo-Electron Microscopy

Micrograph Screening

Arctica with K2 was used for screening grids using Serial Image. Grid screening was done at Pacific Northwest Center for Cryo-Electron Microscopy.

Micrograph Acquisition

Cryo-EM micrographs were imaged on a Titan Krios using a K3 imager. Exposure time was 1.658 with a frame time of 0.0555 at 30 frames per second. Dose was 30 e/A^2. Recorded magnification was 22.5 kX and aperture was set to 100 uM. Images were acquired at the Pacific Northwest Center for Cryo-Electron Microscopy.

Data Processing

Data processing was done on the Boreal cluster housed by Pacific Northwest Center for Cryo-EM. The Relion 3.1 package was used for all processes. Images were imported using 0.514 pixel size and voltage was set to 300 kV. Imported images were motion corrected using a dose per frame of 0.75, Bfactor was set to 150, and number of patches X, Y was set to a value of 5 for both. CTF estimation was conducted by using CTFFINF-4.1 and using the power spectra from motion correction. Gctf was not used. Particles were manually picked using a particle diameter of 500. Image extraction was done without rescaling the particle with the contrast inverted and particles were normalized. 3 rounds of 2D classification were done with the number of classes set to 120, mask diameter set to 550, 25 number of iterations, and other settings were at the default value. 3D initial model set to use 3 classes, mask diameter was 550, symmetry was set to C6, number of iterations was set to 25, number of in-between iterations was set to 100, and number of final iterations was set to 25 with number of pooled particles

set to 500. 3D classification was done with the reference map in absolute grey scale and C6 symmetry. The master diameter was set to 550 and the fast subsets options was turned on. The number of pooled particles was set to 20 and all particles were not read into RAM nor was the process GPU accelerated. 3D auto-refinement was done using C6 symmetry with an initial low-pass filter of 50 and the reference map is not on an absolute greyscale.

Results

P1 does Not Share Synteny of Baseplate and Tail Genes with Mu

The arrangement of the P1 genome has been described as T4 like [96]. Additionally, phage P1 has been identified as having homologs of T4's conserved baseplate core [172]. Furthermore, a simplified core identified in phage Mu and Mu-like phages has been identified in P1 (Figure 19). While the genes that compose T4's conserved core and Mu's simplified core have been identified in P1, synteny between P1 baseplate proteins and that of the simplified core of Mu is not observed (Figure 19B). While the P1 baseplate and tail genes are not clustered as they are in Mu-like phages, genes *bplA*, *pmgA*, and *gp16*, identified in T4's conserved core, are clustered together in the P1 genome (Figure 19B). Together, *bplA*, *pmgA*, and *gp16* encode the P1 baseplate.

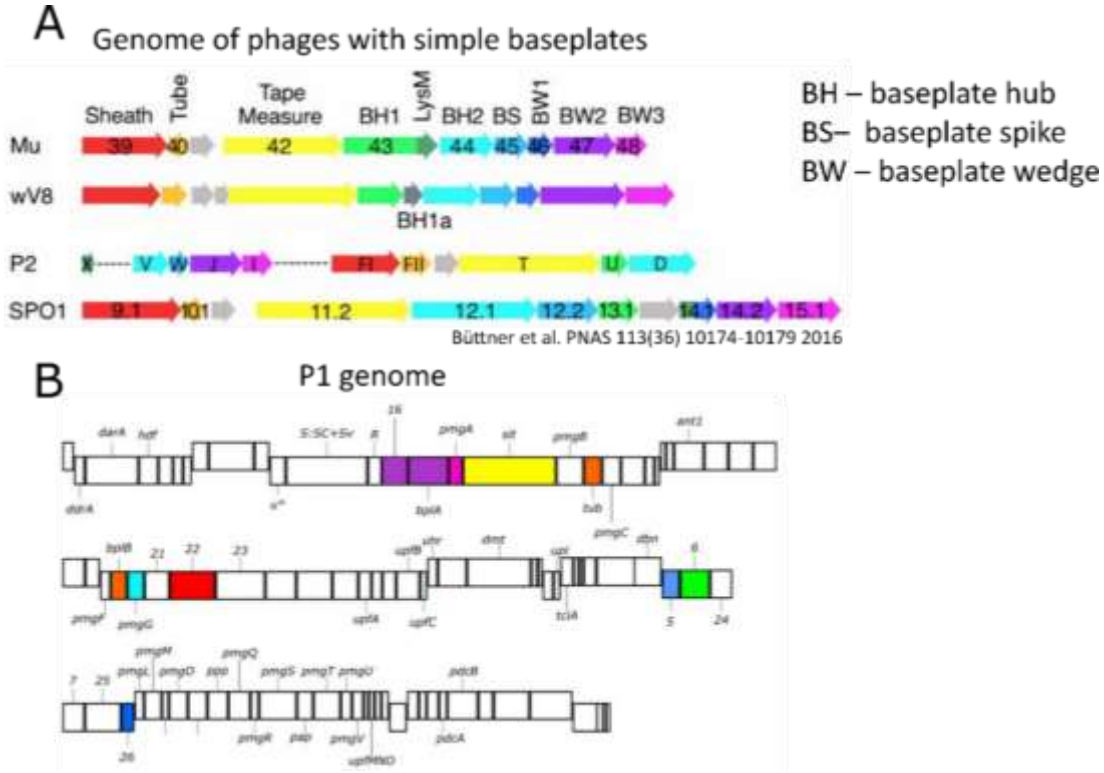

Figure 19 Syntenic comparison of baseplate genes between phages with simple baseplates and P1 baseplate genes. A) Phages with simple baseplates share similar synteny to phage Mu. The genes are color coded to define the function of the gene and corresponding genes have identical colors in different phage background. B) The P1 genome has been separated by open reading frames. Mu homologs identified in the P1 genome and color coded relative to the color scheme of the Mu genome.

Cryo-EM Reveals 3-D Reconstruction of P1

To better understand the structure of the P1 baseplate and tail, cryo-EM was conducted on cesium chloride isopycnic purified P1 virions. To ensure the quality of frozen grids, screening was conducted at a magnification of 75,000 (Figure 20). Screening of the grids revealed coherent grids that would yield 5 days of auto-acquisition. Auto-acquisition yielded 25,475 micrographs. Since the auto-picking software of Relion is not able to auto-pick baseplates from virion micrographs (private

communication with Dr. Petr Leiman), the baseplates were manually picked. Figure 15 illustrates the area of interest that were picked in each micrograph to encompass the P1 baseplate and had a diameter of 500 pixels. Manual picking yielded 11,756 picked particles. Once manual picking was completed, the picked particles were extracted without resizing. The extracted particles were used for 2D classification.

Figure 20 Cryo-EM image of P1 virions during screening of grids. The magnification was 75,000x. This micrograph is a snapshot meant to serve as a sample for the entire grid. Intact P1 virions make up most of the virions identified on the grid, while virions with contracted sheaths, and empty capsids make up a minor subpopulation of virions. The density of virions in this micrograph suggests that the PFU/mL for the sample is good and does not result in overlapping virions nor images and the grid will provide a robust sample size.

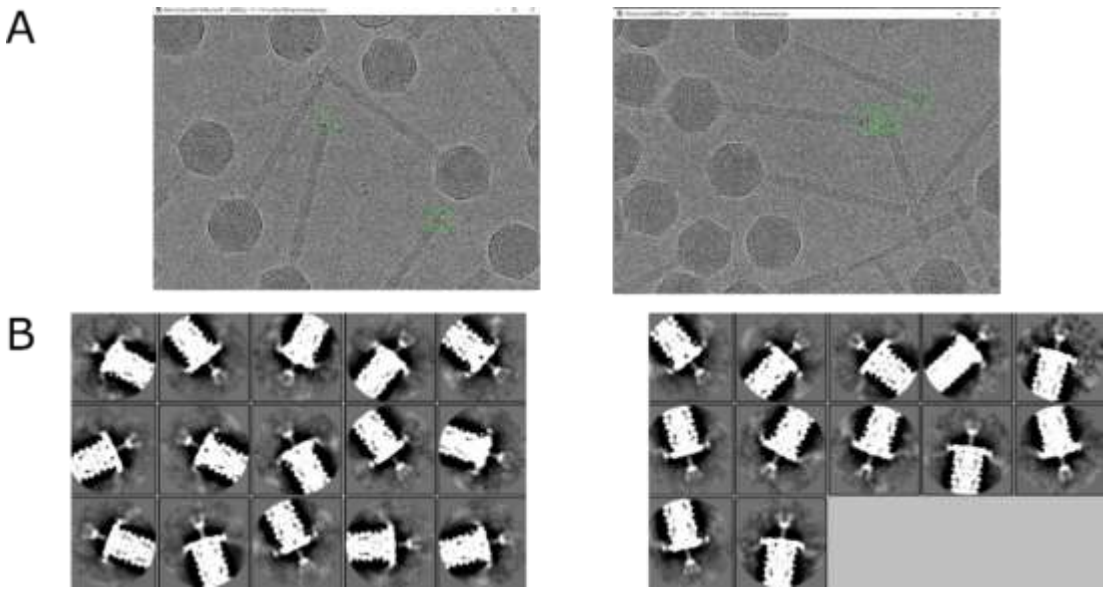

Figure 21 Cryo-Electron micrographs of picked particles and classes selected from 3 successive iterations of 2D classification. A) Two different micrographs of P1 virions with a lime-green circle encompassing the "picked" are used for particles extraction. B) Selected 2D classifications from 3 successive iterations of 2D classification. 27 classes selected from final iteration of 2D classification. These classes were used to generate a 3D model.

3 successive rounds of 2D classification were performed on extracted particles. 10,678 particles remained in the particle pool after unique 2D classes were selected (Figure 21B). Unique 2D classifications were used to ensure complete coverage of the baseplate and the lower tail region of P1. An initial model was constructed using the 37 unique 2D classifications. The initial model, together with the extracted particles were used to make a 3D classification with imposed C6 symmetry. The 3D generated model was then taken through post-processing and the masked 3D density rendered a reconstruction at 3.6 angstrom resolution (Figure 22).

The tail of phage P1 appears to have an outer diameter of 210Å and a 47° rise with a right-handed helical spiral (Figure 22A). The inner diameter of the tail tube is 43Å and appears to be occupied by the tape measure protein Sit (Figure 22B). The uncontracted state of the P1 tail is composed of proteins Tub, PmgG, and Gp22. The P1 sheath has been identified as gp22 through HHpred while the tail tube is composed of Tub and PmgG [172]. The baseplate of P1 has six-fold symmetry and resembles the inner baseplate of T4 with a diameter of 200Å (Figure 23). The P1 plate is composed of Gp26, Gp16, PmgA, and BplA which are each homologous to proteins of the T4 conserved core [172]. The tail fibers could not be resolved, but there are six projections emanating from the baseplate which are suspected to be the N-terminal portions of protein R.

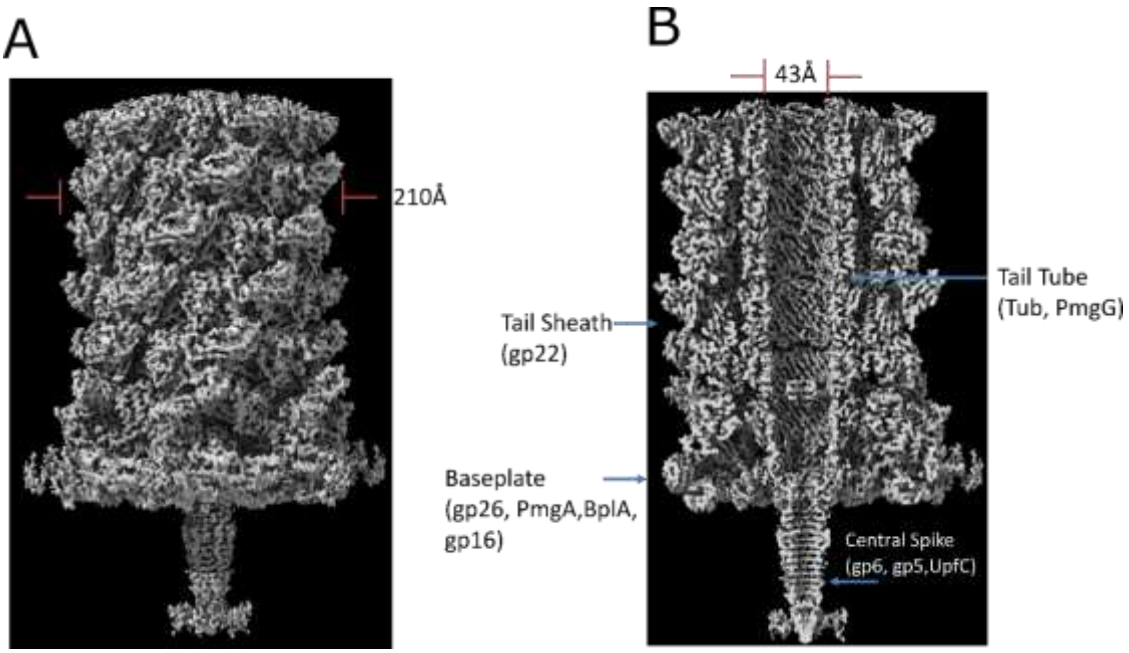

Figure 22 3D density map of the baseplate and lower tail region of P1 at 3.6Å resolution. A) Density of the region displayed in a vertical orientation. B) Lateral cross-section of the baseplate and tail region expose the inner tail tube, thus displaying the diameter of the inner diameter of the tail tube. The blue lines are used to annotate the region of interest with P1 proteins. Measurements were generated in UCSF chimera.

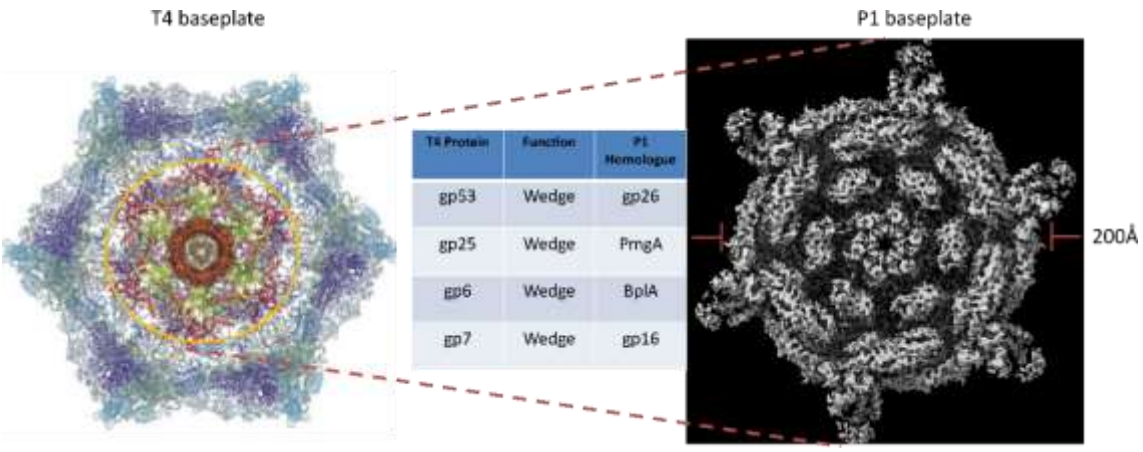

Figure 23 The inner baseplate of T4 resembles the baseplate of P1. The P1 baseplate has T4 homologs displayed in the table between the T4 and P1 baseplates. The orange circle is used to convey the notion that the P1 baseplate only includes proteins within the orange circle. The hashed lines which stretch from the T4 inner base plate to the P1 baseplate are used to illustrate how the P1 baseplate resembles the inner baseplate of T4. The view of the P1 baseplate is a head-on view.

The T4 baseplate is composed of 3 regions, the tail fiber network, the intermediate baseplate, and the inner baseplate. The P1 baseplate resembles that of the T4 inner baseplate. The resemblance of the inner baseplate of T4 is not surprising as P1 homologs of T4's conserved core have been identified in P1 (Table 4). Further, P1 homologs of T4's conserved core can be mapped onto the density of P1 to identify the area where the P1 homologs would occupy (Figure 24). In addition to identifying the areas where P1 homologs would be located in the P1 core, alpha fold prediction was mapped onto the central spike of P1 with atomic precision (Figure 25). While the baseplate and tail proteins of P1 are not clustered together as they are in T4 and Mu-like phages, T4 homologs have been identified in P1 [172].

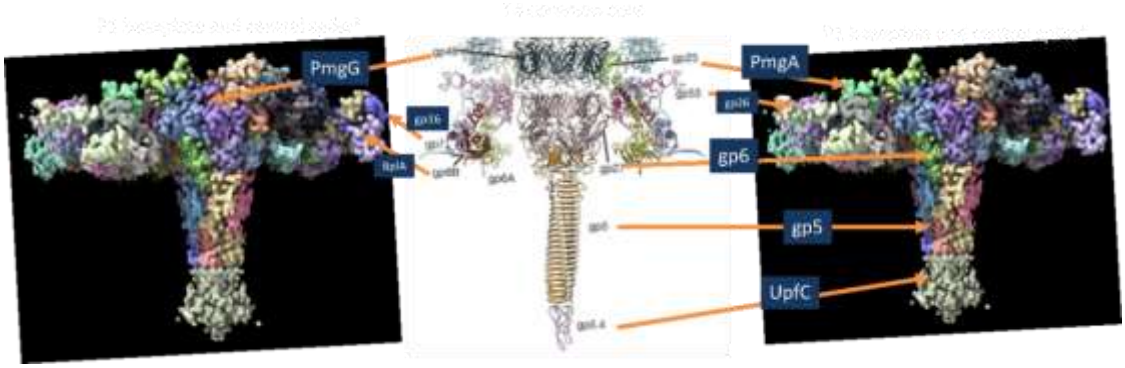

Figure 24 The conserved core of phage T4 is mapped onto the P1. P1 homologs of the T4 conserved core are identified on the right and left images of the segmented P1 density map. The size of the boxes of the P1 homologs are not significant, and the colors of the segmented P1 density map are not significant. Additionally, the P1 density maps are shown in duplicate to decrease crowding of identifying the P1 homologs. The arrows emanating from the T4 conserved core identify regions of the P1 density map where the P1 homologs are likely to be found in the P1 virion.

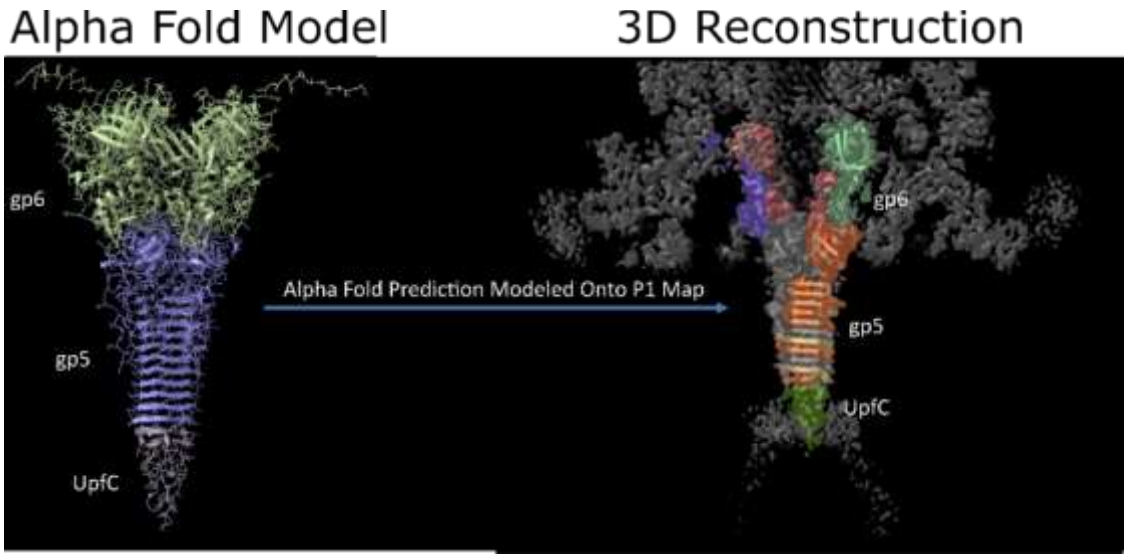

Figure 25 The Alpha fold predicted quaternary structure of the P1 central spike is mapped onto the 3D reconstruction of the P1 baseplate to near perfect alignment. The Alpha fold model predicted the quaternary structure of gp6, gp5, and UpfC to form the central spike of the P1 baseplate. Using Coot, the central spike was mapped onto the 3D density map of the P1 central spike.

Cryo-EM Reveals a Novel Prong on the P1 Central Spike

Cryo-EM of the P1 tail revealed prongs emanating from the tip of the P1 central spike (Figure 26). The prong was evident in 2D classification and 3D refinement revealed the prong at 3.6 A resolution (Figure 26). The symmetry of the prong appears to be six-fold with is congruent with the symmetry of the central spike proteins found in T4 (Figure 26). Due to the limitations of the data set, the asymmetric reconstruction of the P1 baseplate was not able to be rendered. Since the 3D reconstruction has 6-fold symmetry imposed on the P1 baseplate, the prongs appear to be in six copies. The orientations in Figure 20 of the 3D density are meant to convey the orientation of the prongs identified in the 2D classifications which provide a basis of support that the P1 prongs may be found at 6 copies.

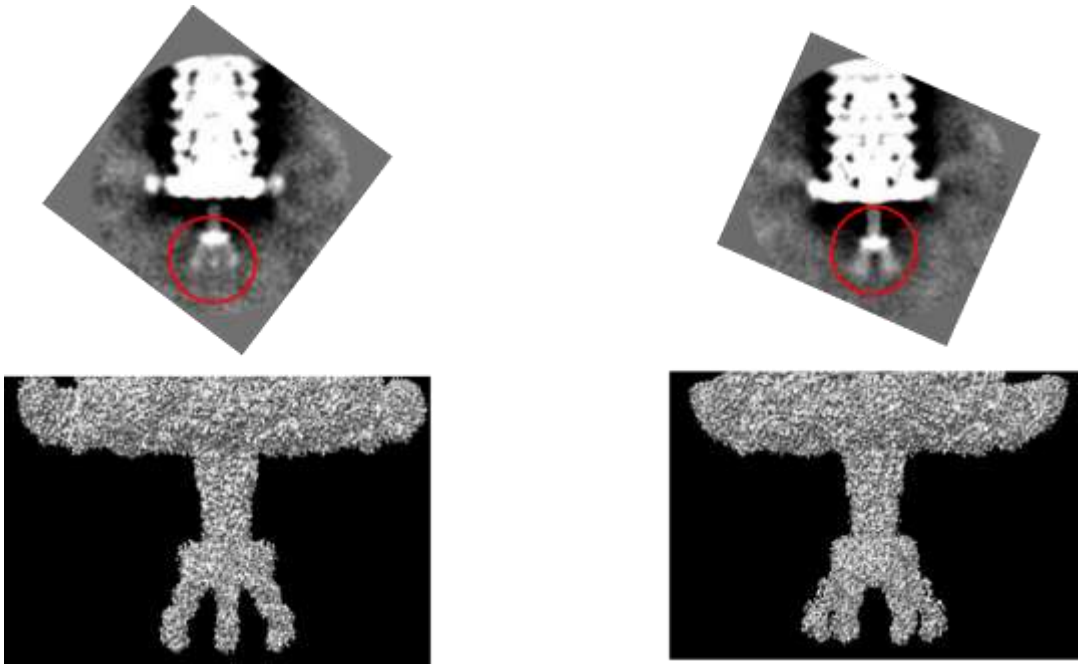

Figure 26 2D classification reveals a prong at the tip of the P1 central spike. The red circle on the 2D classifications is meant to highlight the prong attached to the P1 central spike. The 2D classifications appear above the 3D density in such a way that the orientation of the prong in the 3D density mirrors the orientation of the prong in the 2D classifications.

UpfB Encodes the P1 Central Spike Prong

While the prong has is a unique tail structure, homology is not a useful tool to identify the gene that encodes for the P1 prong. The central spike of P1 has homology to T4 with *upfC* encoding for the tail tip of the central spike of P1. While UpfC is the tail needle of P1 an adjacent gene *pmgB* had no defined function and was found to not be essential [172]. *UpfB* was a prime candidate to be the prong due to its location in the P1 genome and the size of the predicted protein (~26 kDa) (Figure 27). Alpha fold was used to predict the structure of UpfB as a trimer in combination with UpfC (Figure 27). The resulting alpha fold prediction revealed the tail needle with a prong emanating from

UpfC (Figure 27). When applying three-fold symmetry to the alpha fold model, three copies of UpfB are seen emanating from UpfC (Figure 27). While a gene encoding for the prong protein has not been identified in a myophage, there has been only one observation where a prong emanated from the central spike [173] .

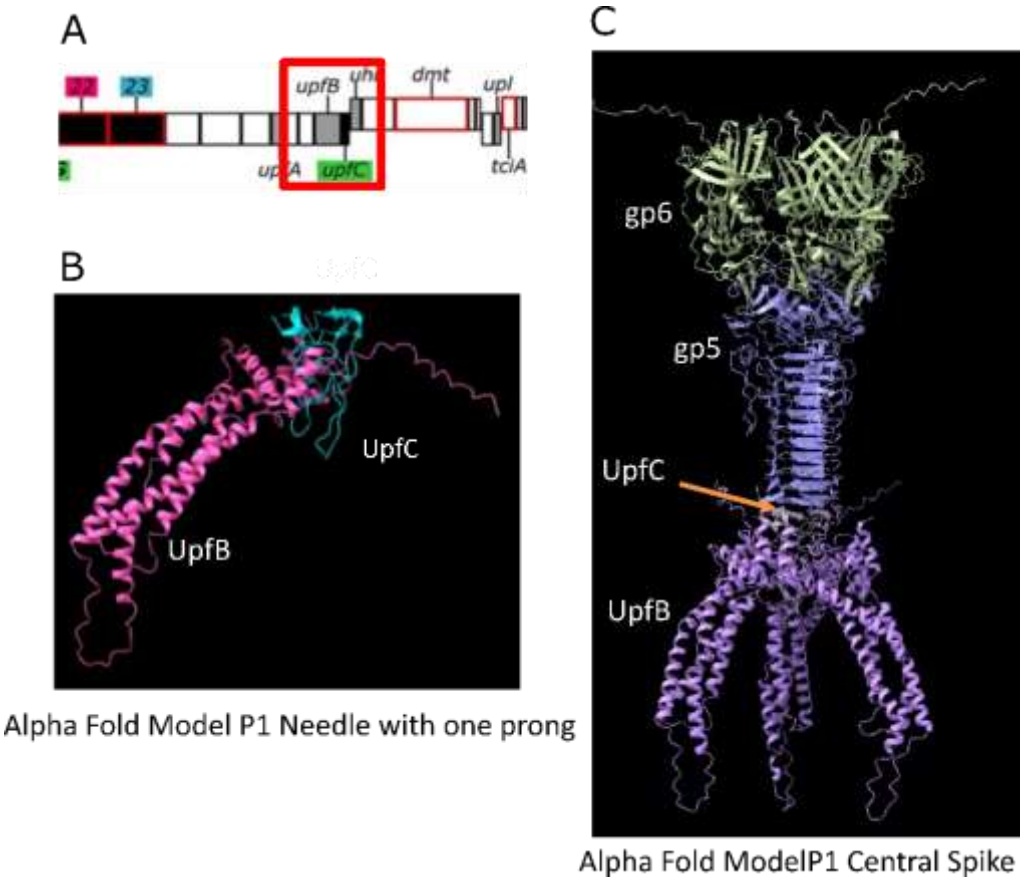

Figure 27 upfB gene location and UpfB Alphafold prediction within different complexes. A) A segment of the P1 genome spanning a region wherein upfB is located. The red box is meant to identify upfB and neighboring genes. B) Alphafold prediction of UpfB and UpfC in complex. One copy of UpfC and UpfB were used for this prediction. C) Alphafold prediction of the P1 central spike. Three copies of UpfB were used, one copy of each UpfC and gp5 were used, and three copies of gp6 were used.

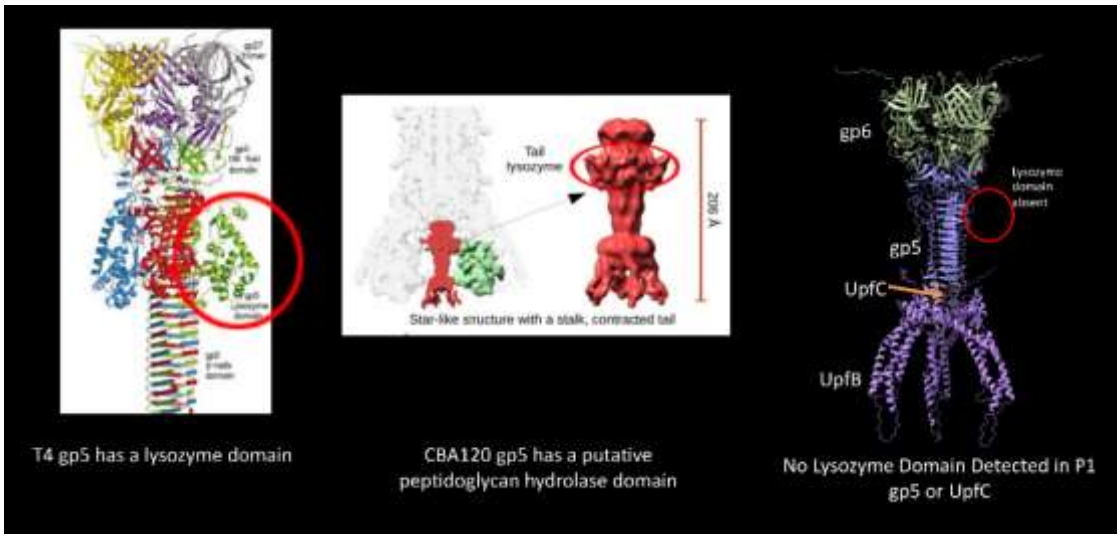

Figure 28 Bioinformatic analysis does not show P1 gp5 or UpfC to have a lysozyme domain. The large red circle encompasses the T4 lysozyme domain located on gp5, while the small red circle on P1 identifies the lack of a lysozyme domain on gp5. The oblong circle on CBA120 identifies a tail lysozyme on the central spike of CBA120. T4 and CBA120 images are credited to Sergey Nazarov.

The central spike of P1 encoded by *gp5* is missing lysozyme domains present in the central spike of T4 and CBA120 (Figure 28). The P1 tape measure protein has transglycolase activity suggested by bioinformatics [172]. Taken together, the cell puncturing mechanism diverges from T4 and is more similar to Mu like phages. The P1 central spike may serve as a membrane puncturing tool but does not serve to degrade the peptidoglycan. Like T4, the central spike of P1 is predicted to detach during infection. However, in the case of P1 this allows for the tape measure protein to make contact with the peptidoglycan in order to degrade the peptidoglycan. Once the peptidoglycan is degraded, the tail tube is able to penetrate the inner membrane of the bacterium thus serving as a conduit to transmit the viral genome (Figure 29).

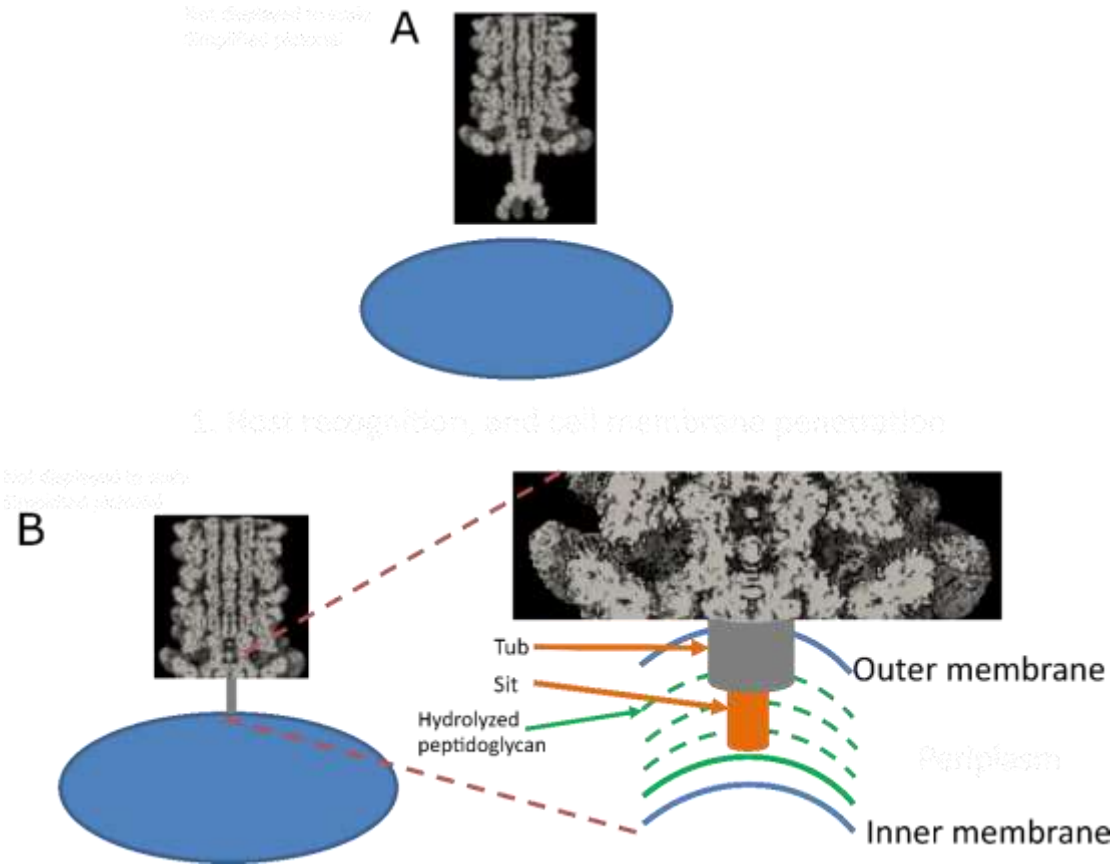

Figure 29 A simplified model of P1 infection on a gram-negative bacterium. A) Depiction of P1 absorbed onto a bacterium. B) Model of the lower region of the P1 tail with an ejected tail tube protein. The picture on the left indicates the attachment of the P1 virion to the cell membrane of a susceptible bacterial host. In this step, the tail prong has punctured the cell membrane and the tail tube is inserted into the bacterial periplasm. The picture on the right is an enlarged image of the section identified on the image of the right. The P1 tail tube and tape measure protein are buried within the bacterial periplasm. The blues curves represent the inner and the outer bacterial membrane and are indicated appropriately. The green curves represent the peptidoglycan. The green curves that are composed of hash marks are peptidoglycan chains that have been hydrolyzed by the P1 tape measure protein while the green curve that is not segmented has not yet been hydrolyzed. Tail fibers are omitted for simplicity.

Infections Assays Reveal Functional Role of UpfB

Infectivity of P1 and *upfB* and *upfC* knockout mutants was determined against *E. coli* strains BL21, WA921, *Serratia marcescens, Klebsiella aerogenes, Salmonella* typ. LT2, *Salmonella* typ. LT2 galE503 (Table 6). Only *E. coli* strains were able to support plaque formation and no significant differences were observed in plaque formation between P1 and P1 Upf knockout mutants (Table 6).

Table 6 Titer of phage P1 backgrounds against various bacterial hosts. The titer was conducted by spotting 10 uL of phage onto a bacterial over lay of each host. A zero indicates that no plaques were observed. Samples with a standard error were performed 3 time. Samples with a value of 0 were performed 3 times. performed 3 time. Phage titer of all phage lysates used were 1x10^^{8} PFU/mL when tittered against BW25113.

	Titer of Phage Background		
Bacteria	**P1**	**Δ*upfB***	**Δ*upfC***
BL21	$4.5x10^6$ +/- $1.0x10^6$	$2.0x10^7$ +/- $9.0x10^6$	$7x10^6$ +/- $2.1x10^6$
WA921	$5.5x10^6$ +/- $2.0x10^6$	$1.0x10^7$ +/- $1.2x10^6$	$7x10^6$+/- $2.1x10^6$
*S. typ.*LT2 galE503	0	0	0
S. typ. LT2	0	0	0
K. aerogenes	0	0	0
S. marcescens	0	0	0

In hosts where plaque formation is not possible, lysogeny formation was measured as a proxy for measuring successful infection events in *S. marcescens*, and *V. cholerae* backgrounds O395, AM19225, N19691, V52, and a strain isolated from the Rio Grande River identified as 4212. *Salmonella enteritidis, Salmonella anatum,* and *Proteus mirabil*is were not found to form chloramphenicol resistant colony forming units after infection with P1 and were not investigated further. Interestingly, *S. marcescens* showed an increase in lysogeny estbalishment for virions lacking UpfB when compared

to wild type P1 virions while virions lacking UpfC were observed to be less infectious when compared to wild type P1 virions (Table 7).

Curiously, lysogeny establishment of *V. cholerae* backgrounds displayed an increase in lysogen formation of P1 mutants over wild type P1 (Table 7). Further, there appeared to be a hierarchical difference between the P1 virion backgrounds when infecting *V. cholerae*. Virions lacking UpfB were more infectious than virions lacking UpfC and wild type virions, while virions lacking UpfC were only more infectious than wild type P1 virions (Table 7). The number of lysogens reported for the *V. cholerae* backgrounds were counted without conducting serial dilutions of the plates which suggesting the difference in the rates of infection between the virion backgrounds are not due to technical errors and, in fact, differences in infection efficiency.

Table 7 Comparative analysis of lysogeny establishment between P1 wild type and unidentified protein function gene knockouts. The fold-change in lysogeny establishment indicate how many more lysogens were established when compared to 1 lysogen formed from the wild type background. As phage titer is not a reasonable way to quantify infection, lysogeny establishment was used as a measure of phage infection. Phage titer of all phage lysates used were 1x10^7 PFU/mL when titered against WA921.

	Fold-Change in Lysogeny Establishment	
Bacteria	**Δ*upfB***	**Δ*upfC***
S. marcescens	4+/-1	5+/-3
Vibrio cholerae		
4212	21+/-8	7+/-2
AM19226	25+/-15	4+/-2
O395	18+/-10	6+/-3
N16961	48+/-28	14+/-6
V52	14+/-2	4+/-0

Conclusion

The P1 genome has been cited as being T4 like [96], but the location of the tail and baseplate proteins suggest otherwise (Figure 19 B). While P1 does have genes homologous to phages T4 and Mu, synteny is not identified in the P1 genome. Interestingly, genes that compose the baseplate are found clustered together (*gp16, bplA, gp26, pmgA*) and maintain synteny with their homologs in T4. Still, genes that compose the P1 central spike, wedges, and the tape measure proteins of P1 along with other essential proteins that compose the P1 baseplate and tail are located in other areas of the P1 genome.

Nearly 10,000 particles remained in the particle pool after 3 successive rounds of 2D classification were completed. The post-processed 3D reconstruction of the P1 lower

tail and baseplate were resolved at a resolution of 3.6 A with imposed 6-fold symmetry. The use of 6-fold symmetry is appropriate due since the canonical symmetry of the T4 baseplate is 6-fold [48]. Additionally, while the baseplate hub of T4 is understood to be composed of a trimer, the plate and the tail sheath are oriented in a 6-fold symmetric geometry, hence the application of 6-fold symmetry to the P1 reconstruction will reveal a 3D reconstruction that is similar to an asymmetric reconstruction. The baseplate of P1 is not as complex as T4 and may resemble the simplicity of the baseplate of phage Mu. In fact, phage Mu encodes 6 structural genes for its' baseplate while P1 encodes 7 structural genes for the P1 baseplate. The baseplate of P1 resembles the inner baseplate of T4 with BplA, gp26, PmgA, and gp16 forming the P1 plate with 6-fold symmetry. P1 genes *bplA, pmgA,* and *gp16* have synteny with T4 genes *gp6, gp25,* and *gp7,* respectively (Figure 19B). While the tube and tail sheath are readily resolved, the density map suggests that the tape measure protein is located in the P1 virion (Figure 22). The addition of the tape measure protein in the P1 virion follows the canonical assembly of myophage tail assembly [48]. All proteins identified in the T4 conserved core have been identified in P1 using HHpred and P1 proteins homologous to the T4 conserved core are mapped onto the P1 3D reconstruction density (Figure 24). This illustrates the conserved nature of the P1 baseplate on a structural perspective, while the location of the genes involved in tail morphogenesis are not syntenic with T4 or phage Mu at large (Figure 19B).

A novel identification revealed by the P1 cryo-EM density map is the prong which emanates from the P1 central spike. A prong has only been identified once before

in phage CBA120, but the resolution of the prong in this report is greater than that reported for CBA120. However, each report presents six copies of the prong emanating from the central spike. While the gene that encodes for the prong was not readily obvious by homology, Alpha Fold revealed a quaternary model of UpfC and UpfB which formed prong (UpfB) attached to the spike tip (UpfC) (Figure 27). UpfC has been identified to have homology to T4's spike tip protein gp5.4 [172]. The placement of *upfB* made it a prime candidate for the P1 prong. Additionally, the placement of *upfB* in the genome is pivotal as *upfC* and *upfB* are both controlled by the late promoter LP_{26} [96]. Interestingly, UpfC has a binding site of the C1 repressor [96]. Presumably, *upfB* and *upfC* are part of the same transcript, thus UpfB is a perfect candidate for the prong protein.

The interesting observations of lysogen establishment between *S. marcescens* and *V. cholerae* pose a unique observation as UpfB knockouts are more infectious than wild type P1 virions when assayed by lysogen formation of *V. cholerae* while no difference in infectivity was observed for *S. marcescens*. Further, virions lacking UpfC appear less infectious than virions lacking UpfB but are more infectious wild type P1 virions. Considering the motility differences between *E. coli* and *V. cholerae*, the addition of the prong may hinder infection as the bacterial locomotive force cause a drag which in turn causes turbulence. In this way, the presence of a prong on P1 virions provides a disadvantage. Although the absence of UpfC in the virion is more infectious than the wild type virions, they are less infectious than virions lacking UpfB. This appears to be a contradiction as the absence of UpfC would also result in an absence of

UpfB from the P1 virion. The high infectivity of UpfC knockouts illustrates the advantage that mutant P1 virions have when infecting *V. cholerae* while also highlighting the significance UpfC serves in infecting *V. cholerae*.

Infection of WA921 did not suffer a dramatic loss of infectivity when UpfC was absent from the P1 virion. Curiously, no difference was observed between the P1 backgrounds when measured by plaque forming units among *E. coli* background WA921. However, Table 6 suggests that P1 virions lacking UpfB were more infectious when infecting *E. coli* strain BL21. Indeed, the difference between titer formation in not extreme, but the data suggests the loss of the P1 tail prong increases infectivity though only marginally.

Since the P1 virion does not require UpfB to infect a range of bacteria, the broad host range may be due to the tail fibers and not the central spike. While the central spike is important for infection, specifically, the presence of UpfB and UpfC each play a role when infecting particular bacterial backgrounds, the broad host range of P1 is a characteristic unrelated to the central spike. Further, P1 is able to infect *S. marcescens* as detected by lysogeny establishment and no change in lysogeny establishment is identified between the three P1 backgrounds reported here. Since all P1 backgrounds in this study are able to infect *S. marcescens* the central spike may not play a role in the broad host range of phage P1. It appears that while the central spike does not play a role in P1's broad host range, the central spike does have a role in infection. Therefore, the differences in the membrane bound proteins between bacterial species that experience

changes in infectivity must be contrasted to specifically identify the characteristics that are overcome - or deleterious in the case of *V. cholerae* – for P1 infection.

CHAPTER V

CONCLUSIONS AND FUTURE WORK

This work has contributed to our understanding of the assembly of the P1 virion by characterizing the viral proteome, identifying phenotypes associated with newly-identified essential genes, determining the high-resolution structure of the P1 baseplate, and examining the packaging and function of the virion-incorporated DarB protein. Essential proteins of the P1 virion have been shown to be involved in specific P1 assembly pathways, and the conserved orthologs of the core baseplate structure could be assigned (Chapter II). Further, the key anti-restriction protein DarB has been genetically interrogated, and helicase and methyltransferase motifs have been observed to play a role in the anti-restriction phenotype against EcoB and EcoK. Curiously, T93A point mutation of the putative methyltransferase active site of DarB did not ablate anti-restriction activity of DarB (Chapter III). Finally, a novel feature of the P1 central spike, the prong, was identified by cryo-electron microscopy and the prong was found to play a role in infection for *V. cholerae* and select *Enterobacteriaceae* (Chapter IV).

Liquid chromatography tandem mass-spectrometry was combined with bioinformatics to identify key P1 structural proteins that were previously identified as the product of a “putative morphogenetic gene” or as having an “unknown protein function”. Further, the 3D reconstruction of the P1 baseplate solidified the notion that the P1 baseplate is simpler than that of phage T4 as the tail fiber network and intermediate baseplate of T4 are absent in phage P1. While this work suggests that

protein R plays a role in connecting the tail fibers to the baseplate, there is not a clear visual representation of the interaction, likely due to the flexibility of this region. The P1 genome has been updated using modern bioinformatic tools and the virion structural proteins have been validated by mass-spectrometry of purified virions (Appendix I).

DarB had already been established as a key protein in P1 the anti-restriction system. Based on previous genetic studies, DarB is believed to protect packaged DNA from EcoB and EcoK restriction activity. The work presented here provides a basis for this to be followed up biochemically. While motifs within the helicase and methyltranferase domains have been shown to be important for DarB activity, biochemical assays which demonstrate the use of AdoMet and ATP by DarB would further suggest a mechanism of anti-restriction activity. Further, the work presented here suggests that protection via methyltransferase activity is dependent on the number of restriction enzyme recognition sequences present within the P1 genome. To explore this notion, the P1 genome can be modified to have more EcoK sites and less EcoB sites. The restriction activity in the presence of an increased or decreased recognition sequence abundance can then be interrogated by DarB methyltransferase mutants. The lack of mCherry to DAPI colocalization in the capsid of P1 virions presented in this work suggests that DNA was able to be ejected from the virion due to auto ejection. On the other hand, the near unit coincidence of DAPI to mCherrry colocalization observed in this work suggests that mCherry is almost always present in the P1 virion when DNA is present. These observations suggest that the DarB located in the P1 virion does not always exit the capsid if DNA is ejected, or perhaps a fraction of this protein remains

with the virion after DNA ejection. Given the conserved domains identified in DarB, it is likely to require both ATP and methyl group donors for its activity, making its most likely site of activity the host cytoplasm during or after DNA ejection. Work in T4, however, has shown that internal head proteins may be able to exercise enzymatic activity even while packaged in the capsid, so the possibility of DarB activity in the capsid cannot be ruled out.

The P1 prong is a novel structure identified only once before in phage CBA120. Here, the prong is presented at 3.6 A resolution. Additionally, P1 mutants lacking the prong had observable differences in infection when infecting *V. cholerae* and *Enterobacteriaceae*. Due to the AlphaFold prediction modeling of UpfB, the gene encoding for the prong is suggested to be *upfB*. It remains to be seen if P1 *upfB* knockouts lack the prong at the central spike when observed using cryo-EM. Additionally, P1 virions lacking *upfB* should be investigated using cryo-EM tomography in an effort to understand the differences in binding mechanics. Further, the true symmetry of the prong is currently unknown. The method used to generate the density map that displays the prong was done so using six-fold symmetry. While the 3D density does mirror the angles that are displayed in 2-D classification, it is reasonable to assume that three-fold symmetry will also mirror the angles presented in 2-D classification. The only method to reconcile the symmetry of the P1 prong as it is on the P1 virion, is to acquire more particles and conduct an asymmetric reconstruction. The particle pool for this analysis is nearly 10,000 particles and imposing six-fold symmetry yields a resolution of 3.6 angstrom. I estimate the amount of information needed to reconstruct a

prong at similar resolution using asymmetric reconstruction would require 50,000 more particles at least which suggests gathering a minimum of 125,000 more micrographs. My estimation is based on private communications with Petr Leiman.

Since P1 has been identified to have a novel prong, and CBA120 has been identified to harbor a prong, it is reasonable to assume other phage species have a prong. The prong may be exclusive to myophages as CBA120 and P1 each have tails that have the capability to contract during infection. Though, unlike CBA120 and T4, P1 does not have a lysozyme domain in its central spike suggesting its mechanism of DNA transmission is similar to phage Mu. The identification of the prong in other phages may simply be a matter of using cryo-EM to conduct additional studies of myophage baseplates. Indeed, NCBI nucleotide Blast reveals a segment of the phage P7 genome to have 99% nucleotide homology to *upfB*. It remains to be seen if cryo-EM reconstruction will reveal a prong in phages that have virtually 100% nucleotide homology to the *upfB* of P1. The evolutionary significance of the prong is currently unknown. Nonetheless, the prong can be examined using optical tweezers or atomic-force microscopy to better understand its binding force to bacterial membrane proteins and even the LPS. It seems intuitive to suggest the prong serves to stabilize the P1 virion during infection, however the role of this structure remains cryptic.

REFERENCES

1. Hendrix, R.W. Bacteriophages: evolution of the majority. *Theor Popul Biol* **2002**, *61*, 471-480, doi:10.1006/tpbi.2002.1590.

2. Brussow, H.; Hendrix, R.W. Phage genomics: small is beautiful. *Cell* **2002**, *108*, 13-16, doi:10.1016/s0092-8674(01)00637-7.

3. Hatfull, G.F.; Hendrix, R.W. Bacteriophages and their genomes. *Curr Opin Virol* **2011**, *1*, 298-303, doi:10.1016/j.coviro.2011.06.009.

4. Casjens, S.R.; Thuman-Commike, P.A. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology* **2011**, *411*, 393-415, doi:10.1016/j.virol.2010.12.046.

5. Bertozzi Silva, J.; Storms, Z.; Sauvageau, D. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* **2016**, *363*, doi:10.1093/femsle/fnw002.

6. Hockenberry, A.J.; Wilke, C.O. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ* **2021**, *9*, e11396, doi:10.7717/peerj.11396.

7. Mackey, M.C.; Santillán, M.; Tyran-Kamińska, M.; Zeron, E.S. The Lysis-Lysogeny Switch. In *Simple Mathematical Models of Gene Regulatory Dynamics*, Mackey, M.C., Santillán, M., Tyran-Kamińska, M., Zeron, E.S., Eds.; Springer International Publishing: Cham, 2016; pp. 99-114.

8. Olszak, T.; Latka, A.; Roszniowski, B.; Valvano, M.A.; Drulis-Kawa, Z. Phage Life Cycles Behind Bacterial Biodiversity. *Curr Med Chem* **2017**, *24*, 3987-4001, doi:10.2174/0929867324666170413100136.

9. Goracci, M.; Pignochino, Y.; Marchio, S. Phage Display-Based Nanotechnology Applications in Cancer Immunotherapy. *Molecules* **2020**, *25*, doi:10.3390/molecules25040843.

10. Ackermann, H.-W.; DuBow, M.S. *Viruses of prokaryotes*; CRC Press: Boca Raton, Fla., 1987.

11. Rakhuba, D.V.; Kolomiets, E.I.; Dey, E.S.; Novik, G.I. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol* **2010**, *59*, 145-155.

12. Blake, C.C.F. Lysozyme. Elliott F. Osserman , Robert E. Canfield , Sherman Beychok. *The Quarterly Review of Biology* **1976**, *51*, 528-528, doi:10.1086/409619.

13. Cahill, J.; Young, R. Phage Lysis: Multiple Genes for Multiple Barriers. *Adv Virus Res* **2019**, *103*, 33-70, doi:10.1016/bs.aivir.2018.09.003.

14. Clokie, M.R.J.; Kropinski, A.M. *Bacteriophages : methods and protocols*; Humana Press: New York, 2009.

15. Wilhelm, S.W.; Suttle, C.A. Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. *BioScience* **1999**, *49*, 781-788, doi:10.2307/1313569.

16. Chaturongakul, S.; Ounjai, P. Phage-host interplay: examples from tailed phages and Gram-negative bacterial pathogens. *Front Microbiol* **2014**, *5*, 442, doi:10.3389/fmicb.2014.00442.

17. Williams, H.T. Phage-induced diversification improves host evolvability. *BMC Evol Biol* **2013**, *13*, 17, doi:10.1186/1471-2148-13-17.

18. Clokie, M.R.; Millard, A.D.; Letarov, A.V.; Heaphy, S. Phages in nature. *Bacteriophage* **2011**, *1*, 31-45, doi:10.4161/bact.1.1.14942.

19. Balcazar, J.L. Bacteriophages as vehicles for antibiotic resistance genes in the environment. *PLoS Pathog* **2014**, *10*, e1004219, doi:10.1371/journal.ppat.1004219.

20. Colomer-Lluch, M.; Jofre, J.; Muniesa, M. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS One* **2011**, *6*, e17549, doi:10.1371/journal.pone.0017549.

21. Wang, X.; Wood, T.K. Cryptic prophages as targets for drug development. *Drug Resist Updat* **2016**, *27*, 30-38, doi:10.1016/j.drup.2016.06.001.

22. Feiner, R.; Argov, T.; Rabinovich, L.; Sigal, N.; Borovok, I.; Herskovits, A.A. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* **2015**, *13*, 641-650, doi:10.1038/nrmicro3527.

23. van Hannen, E.J.; Zwart, G.; van Agterveld, M.P.; Gons, H.J.; Ebert, J.; Laanbroek, H.J. Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses. *Appl Environ Microbiol* **1999**, *65*, 795-801, doi:10.1128/AEM.65.2.795-801.1999.

24. Chapman, M.S.; Liljas, L. Structural folds of viral proteins. *Adv Protein Chem* **2003**, *64*, 125-196, doi:10.1016/s0065-3233(03)01004-0.

25. San Martin, C.; van Raaij, M.J. The so far farthest reaches of the double jelly roll capsid protein fold. *Virol J* **2018**, *15*, 181, doi:10.1186/s12985-018-1097-1.

26. Hendrix, R.W. Evolution: the long evolutionary reach of viruses. *Curr Biol* **1999**, *9*, R914-917, doi:10.1016/s0960-9822(00)80103-7.

27. Baker, M.L.; Jiang, W.; Rixon, F.J.; Chiu, W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol* **2005**, *79*, 14967-14970, doi:10.1128/JVI.79.23.14967-14970.2005.

28. Fokine, A.; Rossmann, M.G. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **2014**, *4*, e28281, doi:10.4161/bact.28281.

29. Ackermann, H.W. 5500 Phages examined in the electron microscope. *Arch Virol* **2007**, *152*, 227-243, doi:10.1007/s00705-006-0849-1.

30. Evilevitch, A.; Lavelle, L.; Knobler, C.M.; Raspaud, E.; Gelbart, W.M. Osmotic pressure inhibition of DNA ejection from phage. *Proc Natl Acad Sci U S A* **2003**, *100*, 9292-9295, doi:10.1073/pnas.1233721100.

31. Chen, D.H.; Baker, M.L.; Hryc, C.F.; DiMaio, F.; Jakana, J.; Wu, W.; Dougherty, M.; Haase-Pettingell, C.; Schmid, M.F.; Jiang, W.; et al. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc Natl Acad Sci U S A* **2011**, *108*, 1355-1360, doi:10.1073/pnas.1015739108.

32. Hendrix, R.W.; Johnson, J.E. Bacteriophage HK97 capsid assembly and maturation. *Adv Exp Med Biol* **2012**, *726*, 351-363, doi:10.1007/978-1-4614-0980-9_15.

33. Leiman, P.G.; Kanamaru, S.; Mesyanzhinov, V.V.; Arisaka, F.; Rossmann, M.G. Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **2003**, *60*, 2356-2370, doi:10.1007/s00018-003-3072-1.

34. Rao, V.B.; Feiss, M. The bacteriophage DNA packaging motor. *Annu Rev Genet* **2008**, *42*, 647-681, doi:10.1146/annurev.genet.42.110807.091545.

35. Black, L.W.; Thomas, J.A. Condensed genome structure. *Adv Exp Med Biol* **2012**, *726*, 469-487, doi:10.1007/978-1-4614-0980-9_21.

36. Mullaney, J.M.; Black, L.W. Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. *J Mol Biol* **1996**, *261*, 372-385, doi:10.1006/jmbi.1996.0470.

37. Rifat, D.; Wright, N.T.; Varney, K.M.; Weber, D.J.; Black, L.W. Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. *J Mol Biol* **2008**, *375*, 720-734, doi:10.1016/j.jmb.2007.10.064.

38. Karam, J.D.; Drake, J.W. *Molecular biology of bacteriophage T4*; American Society for Microbiology: Washington, DC, 1994; pp. xviii, 615 p.

39. Hong, Y.R.; Black, L.W. Protein folding studies in vivo with a bacteriophage T4 expression-packaging-processing vector that delivers encapsidated fusion proteins into bacteria. *Virology* **1993**, *194*, 481-490, doi:10.1006/viro.1993.1287.

40. Guo, F.; Liu, Z.; Vago, F.; Ren, Y.; Wu, W.; Wright, E.T.; Serwer, P.; Jiang, W. Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7. *Proc Natl Acad Sci U S A* **2013**, *110*, 6811-6816, doi:10.1073/pnas.1215563110.

41. Hu, B.; Margolin, W.; Molineux, I.J.; Liu, J. The bacteriophage t7 virion undergoes extensive structural remodeling during infection. *Science* **2013**, *339*, 576-579, doi:10.1126/science.1231887.

42. Choi, K.H.; McPartland, J.; Kaganman, I.; Bowman, V.D.; Rothman-Denes, L.B.; Rossmann, M.G. Insight into DNA and protein transport in double-stranded DNA

viruses: the structure of bacteriophage N4. *J Mol Biol* **2008**, *378*, 726-736, doi:10.1016/j.jmb.2008.02.059.

43. Gleghorn, M.L.; Davydova, E.K.; Rothman-Denes, L.B.; Murakami, K.S. Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* **2008**, *32*, 707-717, doi:10.1016/j.molcel.2008.11.010.

44. Glucksmann, M.A.; Markiewicz, P.; Malone, C.; Rothman-Denes, L.B. Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. *Cell* **1992**, *70*, 491-500, doi:10.1016/0092-8674(92)90173-a.

45. Kazmierczak, K.M.; Davydova, E.K.; Mustaev, A.A.; Rothman-Denes, L.B. The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *EMBO J* **2002**, *21*, 5815-5823, doi:10.1093/emboj/cdf584.

46. Chang, J.; Weigele, P.; King, J.; Chiu, W.; Jiang, W. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* **2006**, *14*, 1073-1082, doi:10.1016/j.str.2006.05.007.

47. Wu, W.; Thomas, J.A.; Cheng, N.; Black, L.W.; Steven, A.C. Bubblegrams reveal the inner body of bacteriophage phiKZ. *Science* **2012**, *335*, 182, doi:10.1126/science.1214120.

48. Leiman, P.G.; Arisaka, F.; van Raaij, M.J.; Kostyuchenko, V.A.; Aksyuk, A.A.; Kanamaru, S.; Rossmann, M.G. Morphogenesis of the T4 tail and tail fibers. *Virol J* **2010**, *7*, 355, doi:10.1186/1743-422X-7-355.

49. Abuladze, N.K.; Gingery, M.; Tsai, J.; Eiserling, F.A. Tail length determination in bacteriophage T4. *Virology* **1994**, *199*, 301-310, doi:10.1006/viro.1994.1128.

50. Pell, L.G.; Liu, A.; Edmonds, L.; Donaldson, L.W.; Howell, P.L.; Davidson, A.R. The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. *J Mol Biol* **2009**, *389*, 938-951, doi:10.1016/j.jmb.2009.04.072.

51. Fokine, A.; Zhang, Z.; Kanamaru, S.; Bowman, V.D.; Aksyuk, A.A.; Arisaka, F.; Rao, V.B.; Rossmann, M.G. The molecular architecture of the bacteriophage T4 neck. *J Mol Biol* **2013**, *425*, 1731-1744, doi:10.1016/j.jmb.2013.02.012.

52. Taylor, N.M.; Prokhorov, N.S.; Guerrero-Ferreira, R.C.; Shneider, M.M.; Browning, C.; Goldie, K.N.; Stahlberg, H.; Leiman, P.G. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **2016**, *533*, 346-352, doi:10.1038/nature17971.

53. Sun, L.; Zhang, X.; Gao, S.; Rao, P.A.; Padilla-Sanchez, V.; Chen, Z.; Sun, S.; Xiang, Y.; Subramaniam, S.; Rao, V.B.; et al. Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat Commun* **2015**, *6*, 7548, doi:10.1038/ncomms8548.

54. Arisaka, F.; Yap, M.L.; Kanamaru, S.; Rossmann, M.G. Molecular assembly and structure of the bacteriophage T4 tail. *Biophys Rev* **2016**, *8*, 385-396, doi:10.1007/s12551-016-0230-x.

55. Ge, H.; Hu, M.; Zhao, G.; Du, Y.; Xu, N.; Chen, X.; Jiao, X. The "fighting wisdom and bravery" of tailed phage and host in the process of adsorption. *Microbiol Res* **2020**, *230*, 126344, doi:10.1016/j.micres.2019.126344.

56. Storms, Z.J.; Arsenault, E.; Sauvageau, D.; Cooper, D.G. Bacteriophage adsorption efficiency and its effect on amplification. *Bioprocess Biosyst Eng* **2010**, *33*, 823-831, doi:10.1007/s00449-009-0405-y.

57. Rossmann, M.G.; Mesyanzhinov, V.V.; Arisaka, F.; Leiman, P.G. The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol* **2004**, *14*, 171-180, doi:10.1016/j.sbi.2004.02.001.

58. Rao, V.B.; Black, L.W. Structure and assembly of bacteriophage T4 head. *Virol J* **2010**, *7*, 356, doi:10.1186/1743-422X-7-356.

59. Egido, J.E.; Costa, A.R.; Aparicio-Maldonado, C.; Haas, P.J.; Brouns, S.J.J. Mechanisms and clinical importance of bacteriophage resistance. *FEMS Microbiol Rev* **2022**, *46*, doi:10.1093/femsre/fuab048.

60. Burmeister, A.R.; Fortier, A.; Roush, C.; Lessing, A.J.; Bender, R.G.; Barahman, R.; Grant, R.; Chan, B.K.; Turner, P.E. Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. *Proc Natl Acad Sci U S A* **2020**, *117*, 11207-11216, doi:10.1073/pnas.1919888117.

61. Gordillo Altamirano, F.; Forsyth, J.H.; Patwa, R.; Kostoulias, X.; Trim, M.; Subedi, D.; Archer, S.K.; Morris, F.C.; Oliveira, C.; Kielty, L.; et al. Bacteriophage-resistant Acinetobacter baumannii are resensitized to antimicrobials. *Nat Microbiol* **2021**, *6*, 157-161, doi:10.1038/s41564-020-00830-7.

62. Trudelle, D.M.; Bryan, D.W.; Hudson, L.K.; Denes, T.G. Cross-resistance to phage infection in Listeria monocytogenes serotype 1/2a mutants. *Food Microbiol* **2019**, *84*, 103239, doi:10.1016/j.fm.2019.06.003.

63. Kortright, K.E.; Chan, B.K.; Turner, P.E. High-throughput discovery of phage receptors using transposon insertion sequencing of bacteria. *Proc Natl Acad Sci U S A* **2020**, *117*, 18670-18679, doi:10.1073/pnas.2001888117.

64. Riede, I.; Eschbach, M.L. Evidence that TraT interacts with OmpA of Escherichia coli. *FEBS Lett* **1986**, *205*, 241-245, doi:10.1016/0014-5793(86)80905-x.

65. McBroom, A.J.; Kuehn, M.J. Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. *Mol Microbiol* **2007**, *63*, 545-558, doi:10.1111/j.1365-2958.2006.05522.x.

66. Manning, A.J.; Kuehn, M.J. Contribution of bacterial outer membrane vesicles to innate bacterial defense. *BMC Microbiol* **2011**, *11*, 258, doi:10.1186/1471-2180-11-258.

67. Testa, S.; Berger, S.; Piccardi, P.; Oechslin, F.; Resch, G.; Mitri, S. Spatial structure affects phage efficacy in infecting dual-strain biofilms of Pseudomonas aeruginosa. *Commun Biol* **2019**, *2*, 405, doi:10.1038/s42003-019-0633-x.

68. Koonin, E.V.; Makarova, K.S.; Wolf, Y.I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol* **2017**, *71*, 233-261, doi:10.1146/annurev-micro-090816-093830.

69. Tock, M.R.; Dryden, D.T. The biology of restriction and anti-restriction. *Curr Opin Microbiol* **2005**, *8*, 466-472, doi:10.1016/j.mib.2005.06.003.

70. Loenen, W.A.; Raleigh, E.A. The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res* **2014**, *42*, 56-69, doi:10.1093/nar/gkt747.

71. Lopatina, A.; Tal, N.; Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu Rev Virol* **2020**, *7*, 371-384, doi:10.1146/annurev-virology-011620-040628.

72. Snyder, L. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol Microbiol* **1995**, *15*, 415-420, doi:10.1111/j.1365-2958.1995.tb02255.x.

73. Chowdhury, R.; Biswas, S.K.; Das, J. Abortive replication of choleraphage phi 149 in Vibrio cholerae biotype el tor. *J Virol* **1989**, *63*, 392-397, doi:10.1128/JVI.63.1.392-397.1989.

74. Fineran, P.C.; Blower, T.R.; Foulds, I.J.; Humphreys, D.P.; Lilley, K.S.; Salmond, G.P. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* **2009**, *106*, 894-899, doi:10.1073/pnas.0808832106.

75. Gerdes, K.; Christensen, S.K.; Lobner-Olesen, A. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* **2005**, *3*, 371-382, doi:10.1038/nrmicro1147.

76. Gerdes, K.; Wagner, E.G. RNA antitoxins. *Curr Opin Microbiol* **2007**, *10*, 117-124, doi:10.1016/j.mib.2007.03.003.

77. Benler, S.; Cobian-Guemes, A.G.; McNair, K.; Hung, S.H.; Levi, K.; Edwards, R.; Rohwer, F. A diversity-generating retroelement encoded by a globally ubiquitous Bacteroides phage. *Microbiome* **2018**, *6*, 191, doi:10.1186/s40168-018-0573-6.

78. Moak, M.; Molineux, I.J. Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol Microbiol* **2004**, *51*, 1169-1183, doi:10.1046/j.1365-2958.2003.03894.x.

79. Rusinov, I.S.; Ershova, A.S.; Karyagina, A.S.; Spirin, S.A.; Alexeevski, A.V. Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses. *BMC Genomics* **2018**, *19*, 885, doi:10.1186/s12864-018-5324-3.

80. Labrie, S.J.; Samson, J.E.; Moineau, S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **2010**, *8*, 317-327, doi:10.1038/nrmicro2315.

81. Otsuka, Y.; Yonesaki, T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Mol Microbiol* **2012**, *83*, 669-681, doi:10.1111/j.1365-2958.2012.07975.x.

82. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol* **1951**, *62*, 293-300, doi:10.1128/jb.62.3.293-300.1951.

83. Bertani, G.; Nice, S.J. Studies on lysogenesis. II. The effect of temperature on the lysogenization of Shigella dysenteriae with phage P1. *J Bacteriol* **1954**, *67*, 202-209, doi:10.1128/jb.67.2.202-209.1954.

84. Lennox, E.S. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1955**, *1*, 190-206, doi:10.1016/0042-6822(55)90016-7.

85. Yarmolinsky, M.B.; Sternberg, N. Bacteriophage P1. In *The Bacteriophages*, Calendar, R., Ed.; Springer US: Boston, MA, 1988; pp. 291-438.

86. Ikeda, H.; Tomizawa, J. Prophage P1, and extrachromosomal replication unit. *Cold Spring Harb Symp Quant Biol* **1968**, *33*, 791-798, doi:10.1101/sqb.1968.033.01.091.

87. Walker, D.H., Jr.; Anderson, T.F. Morphological variants of coliphage P1. *J Virol* **1970**, *5*, 765-782, doi:10.1128/JVI.5.6.765-782.1970.

88. Ikeda, H.; Tomizawa, J.I. Transducing fragments in generalized transduction by phage P1. 3. Studies with small phage particles. *J Mol Biol* **1965**, *14*, 120-129, doi:10.1016/s0022-2836(65)80234-0.

89. Fokine, A.; Leiman, P.G.; Shneider, M.M.; Ahvazi, B.; Boeshans, K.M.; Steven, A.C.; Black, L.W.; Mesyanzhinov, V.V.; Rossmann, M.G. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 7163-7168, doi:10.1073/pnas.0502164102.

90. Kondo, E.; Mitsuhashi, S. Drug Resistance of Enteric Bacteria. Iv. Active Transducing Bacteriophage P1 Cm Produced by the Combination of R Factor with Bacteriophage P1. *J Bacteriol* **1964**, *88*, 1266-1276, doi:10.1128/jb.88.5.1266-1276.1964.

91. Walker, D.H., Jr.; Walker, J.T. Genetic studies of coliphage P1. III. Extended genetic map. *J Virol* **1976**, *20*, 177-187, doi:10.1128/JVI.20.1.177-187.1976.

92. Walker, J.T.; Walker, D.H. Mutations in coliphage p1 affecting host cell lysis. *J Virol* **1980**, *35*, 519-530, doi:10.1128/JVI.35.2.519-530.1980.

93. Walker, J.T.; Walker, D.H., Jr. Coliphage P1 morphogenesis: analysis of mutants by electron microscopy. *J Virol* **1983**, *45*, 1118-1139, doi:10.1128/JVI.45.3.1118-1139.1983.

94. Iida, S.; Streiff, M.B.; Bickle, T.A.; Arber, W. Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology* **1987**, *157*, 156-166, doi:10.1016/0042-6822(87)90324-2.

95. Gill, J.J.; Summer, E.J.; Russell, W.K.; Cologna, S.M.; Carlile, T.M.; Fuller, A.C.; Kitsopoulos, K.; Mebane, L.M.; Parkinson, B.N.; Sullivan, D.; et al. Genomes and characterization of phages Bcep22 and BcepIL02, founders of a novel phage type in Burkholderia cenocepacia. *J Bacteriol* **2011**, *193*, 5300-5313, doi:10.1128/JB.05287-11.

96. Lobocka, M.B.; Rose, D.J.; Plunkett, G., 3rd; Rusin, M.; Samojedny, A.; Lehnherr, H.; Yarmolinsky, M.B.; Blattner, F.R. Genome of bacteriophage P1. *J Bacteriol* **2004**, *186*, 7032-7068, doi:10.1128/JB.186.21.7032-7068.2004.

97. Heinrich, J.; Velleman, M.; Schuster, H. The tripartite immunity system of phages P1 and P7. *FEMS Microbiol Rev* **1995**, *17*, 121-126, doi:10.1111/j.1574-6976.1995.tb00193.x.

98. Biere, A.L.; Citron, M.; Schuster, H. Transcriptional control via translational repression by c4 antisense RNA of bacteriophages P1 and P7. *Genes Dev* **1992**, *6*, 2409-2416, doi:10.1101/gad.6.12a.2409.

99. Rosner, J.L. Formation, induction, and curing of bacteriophage P1 lysogens. *Virology* **1972**, *48*, 679-689, doi:10.1016/0042-6822(72)90152-3.

100. Baumstark, B.R.; Stovall, S.R.; Ashkar, S. Interaction of the P1c1 repressor with P1 DNA: localization of repressor binding sites near the c1 gene. *Virology* **1987**, *156*, 404-413, doi:10.1016/0042-6822(87)90420-x.

101. Velleman, M.; Heinzel, T.; Schuster, H. The Bof protein of bacteriophage P1 exerts its modulating function by formation of a ternary complex with operator DNA and C1 repressor. *J Biol Chem* **1992**, *267*, 12174-12181.

102. Velleman, M.; Heirich, M.; Gunther, A.; Schuster, H. A bacteriophage P1-encoded modulator protein affects the P1 c1 repression system. *J Biol Chem* **1990**, *265*, 18511-18517.

103. Prentki, P.; Chandler, M.; Caro, L. Replication of prophage P1 during the cell cycle of Escherichia coli. *Mol Gen Genet* **1977**, *152*, 71-76, doi:10.1007/BF00264942.

104. Park, K.; Chattoraj, D.K. DnaA boxes in the P1 plasmid origin: the effect of their position on the directionality of replication and plasmid copy number. *J Mol Biol* **2001**, *310*, 69-81, doi:10.1006/jmbi.2001.4741.

105. Wickner, S.; Hoskins, J.; Chattoraj, D.; McKenney, K. Deletion analysis of the mini-P1 plasmid origin of replication and the role of Escherichia coli DnaA protein. *J Biol Chem* **1990**, *265*, 11622-11627.

106. Abeles, A.L.; Reaves, L.D.; Austin, S.J. Protein-DNA interactions in regulation of P1 plasmid replication. *J Bacteriol* **1989**, *171*, 43-52, doi:10.1128/jb.171.1.43-52.1989.

107. Chattoraj, D.K. Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol Microbiol* **2000**, *37*, 467-476, doi:10.1046/j.1365-2958.2000.01986.x.

108. Pal, S.K.; Chattoraj, D.K. P1 plasmid replication: initiator sequestration is inadequate to explain control by initiator-binding sites. *J Bacteriol* **1988**, *170*, 3554-3560, doi:10.1128/jb.170.8.3554-3560.1988.

109. Wickner, S.; Hoskins, J.; McKenney, K. Monomerization of RepA dimers by heat shock proteins activates binding to DNA replication origin. *Proc Natl Acad Sci U S A* **1991**, *88*, 7903-7907, doi:10.1073/pnas.88.18.7903.

110. Abeles, A.L.; Friedman, S.A.; Austin, S.J. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. *J Mol Biol* **1985**, *185*, 261-272, doi:10.1016/0022-2836(85)90402-4.

111. Sternberg, N.; Sauer, B.; Hoess, R.; Abremski, K. Bacteriophage P1 cre gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. *J Mol Biol* **1986**, *187*, 197-212, doi:10.1016/0022-2836(86)90228-7.

112. Magnuson, R.; Lehnherr, H.; Mukhopadhyay, G.; Yarmolinsky, M.B. Autoregulation of the plasmid addiction operon of bacteriophage P1. *J Biol Chem* **1996**, *271*, 18705-18710, doi:10.1074/jbc.271.31.18705.

113. Lehnherr, H.; Yarmolinsky, M.B. Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of Escherichia coli. *Proc Natl Acad Sci U S A* **1995**, *92*, 3274-3277, doi:10.1073/pnas.92.8.3274.

114. Anantharaman, V.; Aravind, L. New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol* **2003**, *4*, R81, doi:10.1186/gb-2003-4-12-r81.

115. Bachi, B.; Reiser, J.; Pirrotta, V. Methylation and cleavage sequences of the EcoP1 restriction-modification enzyme. *J Mol Biol* **1979**, *128*, 143-163, doi:10.1016/0022-2836(79)90123-2.

116. Janscak, P.; Sandmeier, U.; Szczelkun, M.D.; Bickle, T.A. Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J Mol Biol* **2001**, *306*, 417-431, doi:10.1006/jmbi.2000.4411.

117. Devlin, B.H.; Baumstark, B.R.; Scott, J.R. Superimmunity: characterization of a new gene in the immunity region of P1. *Virology* **1982**, *120*, 360-375, doi:10.1016/0042-6822(82)90037-x.

118. Maillou, J.; Dreiseikelmann, B. The sim gene of Escherichia coli phage P1: nucleotide sequence and purification of the processed protein. *Virology* **1990**, *175*, 500-507, doi:10.1016/0042-6822(90)90434-s.

119. Kliem, M.; Dreiseikelmann, B. The superimmunity gene sim of bacteriophage P1 causes superinfection exclusion. *Virology* **1989**, *171*, 350-355, doi:10.1016/0042-6822(89)90602-8.

120. Lu, M.J.; Henning, U. Superinfection exclusion by T-even-type coliphages. *Trends Microbiol* **1994**, *2*, 137-139, doi:10.1016/0966-842x(94)90601-7.

121. Lehnherr, H.; Guidolin, A.; Arber, W. Bacteriophage P1 gene 10 encodes a trans-activating factor required for late gene expression. *J Bacteriol* **1991**, *173*, 6438-6445, doi:10.1128/jb.173.20.6438-6445.1991.

122. Sternberg, N.; Coulby, J. Recognition and cleavage of the bacteriophage P1 packaging site (pac). I. Differential processing of the cleaved ends in vivo. *J Mol Biol* **1987**, *194*, 453-468, doi:10.1016/0022-2836(87)90674-7.

123. Hansen, A.M.; Lehnherr, H.; Wang, X.; Mobley, V.; Jin, D.J. Escherichia coli SspA is a transcription activator for bacteriophage P1 late genes. *Mol Microbiol* **2003**, *48*, 1621-1631, doi:10.1046/j.1365-2958.2003.03533.x.

124. Cohen, G. Electron microscopy study of early lytic replication forms of bacteriophage P1 DNA. *Virology* **1983**, *131*, 159-170, doi:10.1016/0042-6822(83)90542-1.

125. Lehnherr, H.; Guidolin, A.; Arber, W. Mutational analysis of the bacteriophage P1 late promoter sequence Ps. *J Mol Biol* **1992**, *228*, 101-107, doi:10.1016/0022-2836(92)90494-5.

126. Skorupski, K.; Pierce, J.C.; Sauer, B.; Sternberg, N. Bacteriophage P1 genes involved in the recognition and cleavage of the phage packaging site (pac). *J Mol Biol* **1992**, *223*, 977-989, doi:10.1016/0022-2836(92)90256-j.

127. Sternberg, N.; Coulby, J. Cleavage of the bacteriophage P1 packaging site (pac) is regulated by adenine methylation. *Proc Natl Acad Sci U S A* **1990**, *87*, 8070-8074, doi:10.1073/pnas.87.20.8070.

128. Walker, J.T.; Walker, D.H., Jr. Structural proteins of coliphage P1. *Prog Clin Biol Res* **1981**, *64*, 69-77.

129. Piya, D.; Vara, L.; Russell, W.K.; Young, R.; Gill, J.J. The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. *Mol Microbiol* **2017**, *105*, 399-412, doi:10.1111/mmi.13705.

130. Streiff, M.B.; Iida, S.; Bickle, T.A. Expression and proteolytic processing of the darA antirestriction gene product of bacteriophage P1. *Virology* **1987**, *157*, 167-171, doi:10.1016/0042-6822(87)90325-4.

131. Bachi, B.; Arber, W. Physical mapping of BglII, BamHI, EcoRI, HindIII and PstI restriction fragments of bacteriophage P1 DNA. *Mol Gen Genet* **1977**, *153*, 311-324, doi:10.1007/BF00431596.

132. Lin, H.; Rao, V.B.; Black, L.W. Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. *J Mol Biol* **1999**, *289*, 249-260, doi:10.1006/jmbi.1999.2781.

133. Gilcrease, E.B.; Casjens, S.R. The genome sequence of Escherichia coli tailed phage D6 and the diversity of Enterobacteriales circular plasmid prophages. *Virology* **2018**, *515*, 203-214, doi:10.1016/j.virol.2017.12.019.

134. Bai, L.; Wang, J.; Hurley, D.; Yu, Z.; Wang, L.; Chen, Q.; Li, J.; Li, F.; Fanning, S. A novel disrupted mcr-1 gene and a lysogenized phage P1-like sequence detected from a large conjugative plasmid, cultured from a human atypical enteropathogenic Escherichia coli (aEPEC) recovered in China. *J Antimicrob Chemother* **2017**, *72*, 1531-1533, doi:10.1093/jac/dkw564.

135. Rajagopala, S.V.; Casjens, S.; Uetz, P. The protein interaction map of bacteriophage lambda. *BMC Microbiol* **2011**, *11*, 213, doi:10.1186/1471-2180-11-213.

136. Valentine, R.C.; Shapiro, B.M.; Stadtman, E.R. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. *Biochemistry* **1968**, *7*, 2143-2152, doi:10.1021/bi00846a017.

137. Piya, D.K. INTERACTIONS BETWEEN HOST AND PHAGE ENCODED FACTORS SHAPE
PHAGE INFECTION. Texas A&M University, 2018.

138. Keifer, D.Z.; Motwani, T.; Teschke, C.M.; Jarrold, M.F. Measurement of the accurate mass of a 50 MDa infectious virus. *Rapid Commun Mass Spectrom* **2016**, *30*, 1957-1962, doi:10.1002/rcm.7673.

139. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421, doi:10.1186/1471-2105-10-421.

140. Gabler, F.; Nam, S.-Z.; Till, S.; Mirdita, M.; Steinegger, M.; Söding, J.; Lupas, A.N.; Alva, V. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* **2020**, *72*, e108, doi:https://doi.org/10.1002/cpbi.108.

141. Guzman, L.M.; Belin, D.; Carson, M.J.; Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **1995**, *177*, 4121-4130, doi:10.1128/jb.177.14.4121-4130.1995.

142. Zhang, F.; Huang, K.; Yang, X.; Sun, L.; You, J.; Pan, X.; Cui, X.; Yang, H. Characterization of a novel lytic podovirus O4 of Pseudomonas aeruginosa. *Arch Virol* **2018**, *163*, 2377-2383, doi:10.1007/s00705-018-3866-y.

143. Zimmer, M.; Sattelberger, E.; Inman, R.B.; Calendar, R.; Loessner, M.J. Genome and proteome of Listeria monocytogenes phage PSA: an unusual case for programmed + 1 translational frameshifting in structural protein synthesis. *Molecular Microbiology* **2003**, *50*, 303-317, doi:https://doi.org/10.1046/j.1365-2958.2003.03684.x.

144. Roberts, M.D.; Martin, N.L.; Kropinski, A.M. The genome and proteome of coliphage T1. *Virology* **2004**, *318*, 245-266, doi:https://doi.org/10.1016/j.virol.2003.09.020.

145. Huang, R.K.; Khayat, R.; Lee, K.K.; Gertsman, I.; Duda, R.L.; Hendrix, R.W.; Johnson, J.E. The Prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation. *J Mol Biol* **2011**, *408*, 541-554, doi:10.1016/j.jmb.2011.01.016.

146. Effantin, G.; Boulanger, P.; Neumann, E.; Letellier, L.; Conway, J.F. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J Mol Biol* **2006**, *361*, 993-1002, doi:10.1016/j.jmb.2006.06.081.

147. Yap, M.L.; Rossmann, M.G. Structure and function of bacteriophage T4. *Future Microbiol* **2014**, *9*, 1319-1327, doi:10.2217/fmb.14.91.

148. Hendrix, R.W. Bacteriophage HK97: Assembly of the Capsid and Evolutionary Connections. In *Advances in Virus Research*; Academic Press: 2005; Volume 64, pp. 1-14.

149. Roos, W.H.; Gertsman, I.; May, E.R.; Brooks, C.L., 3rd; Johnson, J.E.; Wuite, G.J. Mechanics of bacteriophage maturation. *Proc Natl Acad Sci U S A* **2012**, *109*, 2342-2347, doi:10.1073/pnas.1109590109.

150. Hendrix, R.W. Tail length determination in double-stranded DNA bacteriophages. *Curr Top Microbiol Immunol* **1988**, *136*, 21-29, doi:10.1007/978-3-642-73115-0_2.

151. Lehnherr, H.; Hansen, A.M.; Ilyina, T. Penetration of the bacterial cell wall: a family of lytic transglycosylases in bacteriophages and conjugative plasmids. *Mol Microbiol* **1998**, *30*, 454-457, doi:10.1046/j.1365-2958.1998.01069.x.

152. Xu, J.; Hendrix, R.W.; Duda, R.L. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell* **2004**, *16*, 11-21, doi:10.1016/j.molcel.2004.09.006.

153. Kostyuchenko, V.A.; Navruzbekov, G.A.; Kurochkina, L.P.; Strelkov, S.V.; Mesyanzhinov, V.V.; Rossmann, M.G. The structure of bacteriophage T4 gene product 9: the trigger for tail contraction. *Structure* **1999**, *7*, 1213-1222, doi:10.1016/s0969-2126(00)80055-6.

154. van Raaij, M.J.; Schoehn, G.; Burda, M.R.; Miller, S. Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre11Edited by D. Rees. *Journal of Molecular Biology* **2001**, *314*, 1137-1146, doi:https://doi.org/10.1006/jmbi.2000.5204.

155. Abedon, S.T.; Yin, J. Bacteriophage plaques: theory and analysis. *Methods Mol Biol* **2009**, *501*, 161-174, doi:10.1007/978-1-60327-164-6_17.

156. Zhang, K.; Young, R.; Zeng, L. Bacteriophage P1 does not show spatial preference when infecting Escherichia coli. *Virology* **2020**, *542*, 1-7, doi:10.1016/j.virol.2019.12.012.

157. Luftig, R.B.; Ganz, C. Bacteriophage T4 head morphogenesis. IV. Comparison of gene 16-, 17-, and 49-defective head structures. *J Virol* **1972**, *10*, 545-554, doi:10.1128/JVI.10.3.545-554.1972.

158. Wenzel, S.; Shneider, M.M.; Leiman, P.G.; Kuhn, A.; Kiefer, D. The Central Spike Complex of Bacteriophage T4 Contacts PpiD in the Periplasm of Escherichia coli. *Viruses* **2020**, *12*, doi:10.3390/v12101135.

159. Miller, E.S.; Kutter, E.; Mosig, G.; Arisaka, F.; Kunisawa, T.; Rüger, W. Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **2003**, *67*, 86-156, doi:10.1128/MMBR.67.1.86-156.2003.

160. Duda, R.L.; Oh, B.; Hendrix, R.W. Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation. *J Mol Biol* **2013**, *425*, 2765-2781, doi:10.1016/j.jmb.2013.05.002.

161. Liu, J.; Chen, C.Y.; Shiomi, D.; Niki, H.; Margolin, W. Visualization of bacteriophage P1 infection by cryo-electron tomography of tiny Escherichia coli. *Virology* **2011**, *417*, 304-311, doi:10.1016/j.virol.2011.06.005.

162. Taylor, N.M.I.; van Raaij, M.J.; Leiman, P.G. Contractile injection systems of bacteriophages and related systems. *Mol Microbiol* **2018**, *108*, 6-15, doi:10.1111/mmi.13921.

163. Boulanger, P.; Jacquot, P.; Plançon, L.; Chami, M.; Engel, A.; Parquet, C.; Herbeuval, C.; Letellier, L. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *J Biol Chem* **2008**, *283*, 13556-13564, doi:10.1074/jbc.M800052200.

164. Piuri, M.; Hatfull, G.F. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Mol Microbiol* **2006**, *62*, 1569-1585, doi:10.1111/j.1365-2958.2006.05473.x.

165. Milstone, L.M.; McGuire, J. Different polypeptides form the intermediate filaments in bovine hoof and esophageal epithelium and in aortic endothelium. *J Cell Biol* **1981**, *88*, 312-316, doi:10.1083/jcb.88.2.312.

166. Arber, W.; Dussoix, D. Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. *J Mol Biol* **1962**, *5*, 18-36, doi:10.1016/s0022-2836(62)80058-8.

167. Shao, Q.; Trinh, J.T.; McIntosh, C.S.; Christenson, B.; Balázsi, G.; Zeng, L. Lysis-lysogeny coexistence: prophage integration during lytic development. *Microbiologyopen* **2017**, *6*, doi:10.1002/mbo3.395.

168. Linder, P.; Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* **2011**, *12*, 505-516, doi:10.1038/nrm3154.

169. Bair, C.L.; Black, L.W. A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol* **2007**, *366*, 768-778, doi:10.1016/j.jmb.2006.11.051.

170. Mullaney, J.M.; Black, L.W. Bacteriophage T4 Capsid Packaging and Unpackaging of DNA and Proteins. In *Virus Hybrids as Nanomaterials: Methods and Protocols*, Lin, B., Ratna, B., Eds.; Humana Press: Totowa, NJ, 2014; pp. 69-85.

171. Kaiser, D.; Dworkin, M. Gene transfer to myxobacterium by Escherichia coli phage P1. *Science* **1975**, *187*, 653-654, doi:10.1126/science.803710.

172. Gonzales, M.F.; Piya, D.K.; Koehler, B.; Zhang, K.; Yu, Z.; Zeng, L.; Gill, J.J. New Insights into the Structure and Assembly of Bacteriophage P1. *Viruses* **2022**, *14*, doi:10.3390/v14040678.

173. Nazarov, S. Structure of viral membrane-penetrating machines by electron cryo-microscopy and tomography. Swiss Federal Institute of Technology in Lausanne, 2015.

APPENDIX A

Updated Annotation of the P1 Genome

Bacteriophage P1 has contributed to molecular biology in terms of tools such as the cre-lox molecular engineering system in addition to enabling the mapping of the *E. coli* genome [1][2]. While the legacy of P1 lives on, modern bioinformatic tools have not been applied to the P1 functional annotation. The current P1 genome is largely ambiguous with nearly half of the genes having no known function. While an amber library has elucidated essential P1 genes, there remains a large set of genes that have no functional annotation. Recent advancements in bioinformatics have enabled a comparative analysis of a protein sequence to be aligned with every protein sequence deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Base (PBD) [3]. HHpred is a powerful bioinformatic tool that uses the primary sequence of an amino acid to determine the probability and identity of closely matched amino acid sequence deposited in PDB using an all-against-all approach [4]. In this way, query protein sequences can be matched to well-studied proteins and a functional role can be assigned. HHpred provides the probability, coverage, and identity of query proteins as they relate to the homologs in PDB. Using HHpred, each CDS within the P1 proteome has been matched by homology to an existing entry in PDB and reported in the updated CDS notes only if the probability of the target protein is 80% or higher. The total amino acids from the P1 query are reported, along with the aligned segment of the P1 query protein and the aligned segment of their relative homologue. Taken together, a specific stretch of a P1 amino acid sequence is reported as having a high probability of

having the functional role as its identified homologue within PDB. The PDB identity is noted in the updated annotation. In addition, the P1 CDS is also updated to include proteins that have been identified as structural. The structural presence of P1 proteins have been investigated after the P1 genome was initially deposited. The laboratories of Małagorzata Lobocka and Jason J. Gill have each conducted separate purification and sequencing methods of the P1 virion. The independent investigations from these labs have been used to annotate pertinent P1 CDS as being structural. The methods that have identified a P1 CDS in their workflow have been added to the CDS notes and the appropriate PMID has been added in cases where structural proteins have been reported in peer-reviewed literature. Finally, a color-coded genome map of phage P1 has been provided to supply a visual representation of the structural proteins of P1.

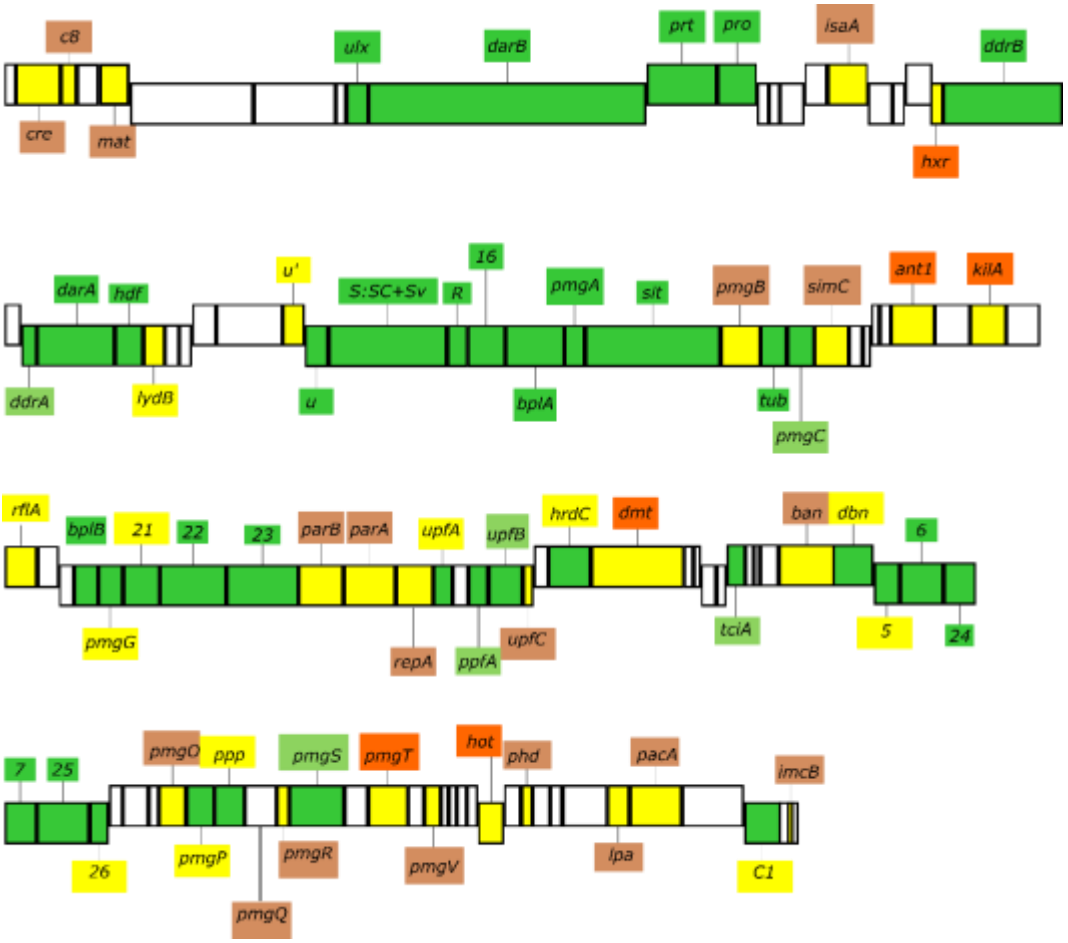

Figure 30 The P1 genomic map colored according to the structural role of each gene. The color gene corresponds to the structural role of the gene with green being structural, yellow being possible, and white having no structural role. The color of the gene name corresponds to the confidence the encoded protein is structural. Green notes that both are confident, light green notes that one investigator is confident while the other has only observed the protein. Orange notes that both investigators have observed the protein in their work, but are not confident the protein is structural. Brown notes that only one investigator has observed the protein in their work and are not confident it is structural.

Table 8 Update for the NCBI Genbank CDS Comment for each P1 protein. HHPred was used to align each P1 protein to Protein Data Bank. Only matches to P1 proteins that received a probability value of 80% or higher were added "CDS comment update". Further, the status of each gene as structural as they are identified in Figure 24 is added to the CDS comment update. Additionally, the methods used to identify structural proteins have been described in the CDS comment update. The entries in "Note first clause" are the current notes for each gene which describes the function of the open reading frame. CDS comment updates are blank for the genes.

Name	Note first clause	DBxrefs	CDS Comment Update
cra	cre associated function	GeneID:2777485	
YP_006471.1	None	Genbank:YP_006471.1	No update necessary
cre	cyclization recombinase	GeneID:2777477	
YP_006472.1	None	Genbank:YP_006472.1	Cre has been identified in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. HHpred identified the aa sequence as having a probability of 99.91 of being Recombinase cre of Enterobacteria phage P1 (PDB entry: 3MGV_B). CDS has 100% identity to 343 aa sequence of phage P1. Additionally, HHpred identified the aa sequence as having a probability of 99.03 of being site-specific recombinase xerd of Escherichia coli (PDB entry: 1A0P). CDS has 15% identity to xerd with 266 total aa spanning 34...333 aligned to aa 16...286 of xerd. Other bacterial DNA recombinases have also been identified.
c8	None	GeneID:2777468	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006473.1	None	Genbank:YP_006473.1	C8 has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification.
ref	recombination enhancement function	GeneID:2777429	
YP_006474.1	None	Genbank:YP_006474.1	HHpred identified the aa sequence as having a probability of 100.00 of being Recombination enhancement function protein of Enterobacteria phage P1 (PDB entry: 3PLW_A). CDS has 100% identity to 186 aa sequence of phage P1.
mat	maturation control, corresponds to former gene 1	GeneID:2777380	
YP_006475.1	None	Genbank:YP_006475.1	Mat has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification. In addition, mat has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Mat may be a structural protein. HHpred does not provide an alignment to PDB at a probability above 80%.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
res	restriction component of host specificity DNA complex (hsdR) in type III restriction-modification enzyme EcoP1	GeneID:2777489	
YP_006476.1	None	Genbank:YP_006476.1	HHpred identified the aa sequence as having a probability of 100.00 of being EcoP15I of Escherichia coli (PDB entry: 4ZCF_C). CDS has 92% identity to EcoP15I with 970 total aa spanning 1...970 aligned to aa 1...970 of EcoP15I. Other bacterial modification systems have also been identified.
mod	mod; N6-adenine methyltransferase gene	GeneID:2777490	
lxc	lowers expression of c1; modulator of C1 action	GeneID:2777417	
YP_006477.1	None	Genbank:YP_006477.1	HHpred does not provide an alignment to PDB at a probability above 80%.
ulx	gene upstream of lxc	GeneID:2777418	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006478.1	None	Genbank:YP_006478.1	UlX enables the recruitment of DarB to the P1 virion. P1 lysogens lacking *ulx* have decreased DarB associated with the P1 virion and increased sensitivity to EcoB and EcoK when compared to wild-type. Ulx is a structural protein. PMID: 28509398
darB	defense against restriction	GeneID:2777481	
YP_006479.1	None	Genbank:YP_006479.1	DarB confers anti-restriction activity against EcoB and EcoK endonucleases. P1 virions lacking *darB* have increased sensitivity to EcoB and EcoK when compared to wild-type lysogens when grown in a modification deficient host. DarB is a structural protein. PMID: 28509398
prt	portal, putative	GeneID:2777482	
YP_006480.1	None	Genbank:YP_006480.1	HHpred identified the aa sequence as having a probability of 100.00 of being gp20 of Enterobacteria phage T4 (PDB entry: 3JA7_J). CDS has 15% identity to gp20 with 423 total aa spanning 89...553 aligned to aa 5...451 of gp20. Prt is a structural protein. PMID: 35458408
pro	head processing	GeneID:2777381	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006481.1	None	Genbank:YP_006481.1	HHpred identified the aa sequence as having a probability of 87.18 of being Prohead core protein protease of Enterobacteria phage T4 (PDB entry: 5JBL_C). CDS has 18% identity to prohead core protein protease with 131 total aa spanning 19...176 aligned to aa 67...231 of prohead core protein protease. Pro is a structural protein. PMID: 35458408
lydE	lysis determinant, putative antiholin	GeneID:2777465	
YP_006482.1	None	Genbank:YP_006482.1	HHpred does not provide an alignment to PDB at a probability above 80%.
lydD	lysis determinant, putative holin	GeneID:2777382	
YP_006483.1	None	Genbank:YP_006483.1	HHpred does not provide an alignment to PDB at a probability above 80%.
lyz	lysozyme, former name: gene 17	GeneID:2777454	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006484.1	None	Genbank:YP_006484.1	HHpred identified the aa sequence as having a probability of 99.95 of being Lysozyme (E.C.3.2.1.17) of Enterobacteria phage p1 (PDB entry: 1XJT_A). CDS has 100% identity to lysozyme with 185 total aa spanning 1...185 aligned to aa 1...185 of lysozyme. Other bacterial and phage lysozymes were identified at above 99% probability.
ssb	gene for the single-stranded DNA binding protein homolog	GeneID:2777409	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006485.1	None	Genbank:YP_006485.1	HHpred identified the aa sequence as having a probability of 99.95 of being Single stranded DNA binding protein/RNA of Escherichia coli (PDB entry: 1EQQ_B). CDS has 70% identity to single stranded DNA binding protein/RNA with 162 total aa spanning 1...162 aligned to aa 1...178 of single stranded DNA binding protein/RNA. Other bacterial and single stranded DNA binding proteins were identified at above 99% probability.
isaA	IS1 insertion-associated gene	GeneID:2777410	
YP_006486.1	None	Genbank:YP_006486.1	IsaA has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification.
insB	IS1 transposition protein InsB	GeneID:2777463	
YP_006487.1	None	Genbank:YP_006487.1	HHpred does not provide an alignment to PDB at a probability above 80%.
insA	IS1 transposition protein	GeneID:2777464	
YP_006488.1	None	Genbank:YP_006488.1	HHpred identified the aa sequence as having a probability of 94.37 of being uncharacterized HTH-type transcriptional regulator ygiT of Escherichia coli (PDB entry: 2KZ8_A). CDS has 21% identity to HTH-type transcriptional regulator ygiT with 80 total aa spanning 7...86 aligned to aa 2...105 of HTH-type transcriptional regulator ygiT. Other bacterial and transcriptional factors were identified at greater than 80% probability.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
isaB	IS1 insertion-associated gene	GeneID:2777411	
YP_006489.1	None	Genbank:YP_006489.1	Hhpred does not provide an alignment to PDB at a probability above 80%.
hxr	homolog of xre	GeneID:2777412	
YP_006490.1	None	Genbank:YP_006490.1	Hxr has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification. In addition, hxr has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Hxr may be a structural protein. HHpred probability below 80%
ddrB	gene downstream of darA	GeneID:2777413	
YP_006491.1	None	Genbank:YP_006491.1	DdrB part of the multi-component anti-restriction of phage P1. DdrB is a negative regulator of DarB. The absence of ddrB results in an increased accumulation of DarB in the P1 virion. DarB is a structural protein. PMID: 28509398
iddB	gene internal to ddrB	GeneID:2777466	
YP_006492.1	None	Genbank:YP_006492.1	Hhpred does not provide an alignment to PDB at a probability above 80%.
ddrA	gene downstream of darA	GeneID:2777414	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006493.1	None	Genbank:YP_006493.1	DdrA part of the multi-component anti-restriction of phage P1. DdrA displays anti-restriction activity against EcoA. Additionally, ddrA also recruits ddrB, DarB, and ulx to the P1 virion. DdrA is a structural protein. PMID: 28509398
darA	defense against restriction	GeneID:2777415	
YP_006494.1	None	Genbank:YP_006494.1	DarA part of the multi-component anti-restriction of phage P1. DarA, together with hdf, is responsible for recruiting hdf and ddrA to the P1 virion. Additionally, the presence of DarA influences the capsid assembly of P1 to predominantly favor capsids with a triangulation number of 13 over capsids with a triangulation number of 7. Further, the absence of DarA results in the P1 capsid assembly pathway to favor capsids with a triangulation number of 7 over capsids with a triangulation number of 13. DarA is a structural protein. PMID: 28509398
hdf	homolog of darA fragment	GeneID:2777469	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006495.1	None	Genbank:YP_006495.1	Hdf part of the multi-component anti-restriction of phage P1. Hdf, together with DarA, is responsible for recruiting ddrA to the P1 virion. Additionally, the presence of DarA influences the capsid assembly of P1 to predominantly favor capsids with a triangulation number of 13 over capsids with a triangulation number of 7. Further, the absence of DarA results in the P1 capsid assembly pathway to favor capsids with a triangulation number of 7 over capsids with a triangulation number of 13. DarA is a structural protein. PMID: 28509398
lydB	lysis determinant, prevents premature lysis	GeneID:2777470	
YP_006496.1	None	Genbank:YP_006496.1	LydB has been observed by whole phage shotgun analysis using HPLC purified bands as well as ammonium acetate triple purification. LydB is a structural protein. HHpred identified matched proteins to be at a probability below 80%

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
lydA	lysis determinant; holin	GeneID:2777471	
YP_006497.1	None	Genbank:YP_006497.1	Hhpred does not provide an alignment to PDB at a probability above 80%.
lydC	lysis determinant	GeneID:2777416	
YP_006498.1	None	Genbank:YP_006498.1	Hhpred does not provide an alignment to PDB at a probability above 80%.
cin	site-specific recombinase, DNA invertase	GeneID:2777388	
YP_006499.1	None	Genbank:YP_006499.1	HHpred identified the aa sequence as having a probability of 99.91 of being DNA-invertase of Enterobacteria phage Mu (PDB entry: 3UJ3_X). CDS has 76% identity to DNA-invertase with 180 total aa spanning 1...180 aligned to aa 1...180 of DNA-invertase. Other bacterial DNA-invertase were identified at greater than 80% probability.
Sv prime	None	GeneID:2777492	
U prime	encodes assembly or structural protein gpU prime of tail fibers in C(-) phage	GeneID:2777419	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006500.1	None	Genbank:YP_006500.1	HHpred identified the aa sequence as having a probability of 99.93 of being protein U of Enterobacteria phage Mu (PDB entry: 5YVQ_B). CDS has 56% identity to protein U with 175 total aa spanning 1...177 aligned to aa 1...175 of protein U. Protein U prime is structural. PMID: 35458408
U	encodes assembly or structural protein gpU of tail fibers in C(+) phage	GeneID:2777436	
YP_006501.1	None	Genbank:YP_006501.1	HHpred identified the aa sequence as having a probability of 100.00 of being protein U of Enterobacteria phage Mu (PDB entry: 5YVQ_B). CDS has 56% identity to protein U with 175 total aa spanning 1...175 aligned to aa 1...175 of protein U. U has been observed by whole phage shotgun analysis using HPLC purified bands in addition to ammonium acetate triple purification. U is a structural protein.
S	encodes structural protein of tail fibers gpS in C(+) phage	GeneID:2777437	
YP_006502.1	None	Genbank:YP_006502.1	HHpred identified the aa sequence as having a probability of 100.00 of being protein S of Enterobacteria phage Mu (PDB entry: 5YVQ_A). CDS has 86% identity to protein S with 337 total aa spanning 651...987 aligned to aa 165...504 of protein S. Protein S is structural. PMID: 35458408

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
R	tail fiber structure or assembly	GeneID:2777404	
YP_006503.1	None	Genbank:YP_006503.1	HHpred identified the aa sequence as having a probability of 98.72 of being protein gp105 of Listeria phage A511 Mu (PDB entry: 6HHK_C). Protein R is structural. PMID: 35458408
16	baseplate or tail tube	GeneID:2777405	
YP_006504.1	None	Genbank:YP_006504.1	HHpred does not provide an alignment to PDB at a probability above 80%.
bplA	baseplate structure; may correspond to gene 3	GeneID:2777406	
YP_006505.1	None	Genbank:YP_006505.1	HHpred identified the aa sequence as having a probability of 99.47 of being Baseplate wedge protein gp7 of Enterobacteria phage T4 (PDB entry: 5HX2_E). CDS has 20% identity to Baseplate wedge protein gp7 with 307 total aa spanning 5...319 aligned to aa 23...366 of Baseplate wedge protein gp7. HHpred identified the aa sequence as having a probability of 99.11 of being Baseplate wedge protein gp6 of Enterobacteria phage T4 (PDB entry: 5IV5_EB). CDS has 18% identity to Baseplate wedge protein gp6 with 384 total aa spanning 5...408 aligned to aa 23...448 of Baseplate wedge protein gp6. BplA is structural. PMID: 35458408
pmgA	putative morphogenetic function	GeneID:2777407	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006506.1	None	Genbank:YP_006506.1	HHpred identified the aa sequence as having a probability of 99.4 of being protein gp25 of Enterobacteria phage T4 (PDB entry: 5IW9_B). Lysogens lacking *pmgA* produce amorphous capsids which appear to lack tails. PmgA is structural. PMID: 35458408
sit	structural injection transglycosylase; by homology	GeneID:2777408	
YP_006507.1	None	Genbank:YP_006507.1	HHpred identified the aa sequence as having a probability of 98.92 of being soluble lytic murein transglycosylase of Pseudomonas aeruginosa (PDB entry: 5OHU_A). Sit is also the P1 tape measure protein. Sit is structural. PMID: 35458408
pmgB	putative morphogenetic function	GeneID:2777478	
YP_006508.1	None	Genbank:YP_006508.1	PmgB has been identified in low abundance by whole phage shotgun analysis by ammonium acetate triple purification. PmgB is an essential protein though its function is not known. Lysogens lacking *pmgB produce* amorphous capsids which appear to lack tails. PMID: 35458408
tub	tail tube structure	GeneID:2777479	
YP_006509.1	None	Genbank:YP_006509.1	HHpred identified the aa sequence as having a probability of 99.94 of being tail tube protein gp19 of Enterobacteria phage T4 (PDB entry: 5IV5_IB). Tub is structural. PMID: 35458408
pmgC	putative morphogenetic function	GeneID:2777480	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006510.1	None	Genbank:YP_006510.1	HHpred identified the aa sequence as having a probability of 99.06 of being gp11 tail adaptor of Enterobacteria phage T7 (PDB entry: 7BOX_N). Lysogens lacking *pmgC* produce amorphous capsids as well as tails, yet capsids and tails are not connected. PmgC is structural. PMID: 35458408
simC	superimmunity; confers superimmunity when in high copy number	GeneID:2777491	
YP_006511.1	None	Genbank:YP_006511.1	SimC has been identified in low abundance by whole phage shotgun analysis by ammonium acetate triple purification. HHpred identified the aa sequence as having a probability of 94.32 of being putative nucleic acid-binding lipoprotein of Klebsiella pneumoniae subsp. pneumoniae MGH 78578 (PDB entry: 3F1Z_J). CDS has 16% identity to putative nucleic acid-binding lipoprotein with 94 total aa spanning 65...160 aligned to aa 29...130 of putative nucleic acid-binding lipoprotein.
simB	superimmunity linked function	GeneID:2777487	
YP_006512.1	None	Genbank:YP_006512.1	HHpred does not provide an alignment to PDB at a probability above 80%.
simA	superimmunity linked function	GeneID:2777488	
YP_006513.1	None	Genbank:YP_006513.1	HHpred does not provide an alignment to PDB at a probability above 80%.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
c4	c4 RNA	GeneID:2777495	
icd	causes reversible inhibition of cell division	GeneID:2777450	
YP_006514.1	None	Genbank:YP_006514.1	HHpred does not provide an alignment to PDB at a probability above 80%.
ant1	antagonism of C1 repression	GeneID:2777451	
YP_006515.1	None	Genbank:YP_006515.1	Ant1 has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification. In addition, ant1 has been identified in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Ant1 may be a structural protein. HHpred probability below 80%
ant2	antagonism of C1 repression, contained within ant1 and translated in the same frame	GeneID:2777452	
YP_006516.1	None	Genbank:YP_006516.1	HHpred does not provide an alignment to PDB at a probability above 80%.
ask	None	GeneID:2777453	
kilA	expression can kill host	GeneID:2777383	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006517.1	None	Genbank:YP_006517.1	KilA has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification. HHpred probability below 80%
repL	lytic replication	GeneID:2777395	
YP_006518.1	None	Genbank:YP_006518.1	HHpred identified the aa sequence as having a probability of 95.58 of being DNA replication protein DNAD of Bacillus subtilis (PDB entry: 2V79_A). CDS has 13% identity to DNA replication protein DNAD with 85 total aa spanning 30...118 aligned to aa 28...112 of DNA replication protein DNAD.
rlfA	replication linked function	GeneID:2777435	
YP_006519.1	None	Genbank:YP_006519.1	RflA has been observed by whole phage shotgun analysis using HPLC purified bands, ammonium acetate triple purification, as well as double cesium chloride purification. RflA may be a structural protein. HHpred does not provide an alignment to PDB at a probability above 80%.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
rlfB	replication linked function	GeneID:2777476	
YP_006520.1	None	Genbank:YP_006520.1	HHpred does not provide an alignment to PDB at a probability above 80%.
pmgF	putative morphogenetic function	GeneID:2777443	
YP_006521.1	None	Genbank:YP_006521.1	HHpred does not provide an alignment to PDB at a probability above 80%.
bplB	baseplate structure, by homology	GeneID:2777444	
YP_006522.1	None	Genbank:YP_006522.1	HHpred identified the aa sequence as having a probability of 99.79 of being gp19 tail tube protein of Enterobacteria phage T4 (PDB entry: 5IV5_IB). PmgB is structural. PMID: 35458408
pmgG	putative morphogenetic function	GeneID:2777384	
YP_006523.1	None	Genbank:YP_006523.1	HHpred identified the aa sequence as having a probability of 97.3 of being gp48 tail tube assembly protein of Enterobacteria phage T4 (PDB entry: 5IV5_R). PmgG is structural. PMID: 35458408
21	baseplate or tail tube	GeneID:2777385	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006524.1	None	Genbank:YP_006524.1	Gp21 has been observed in a moderate amount by whole phage shotgun analysis by ammonium acetate triple purification. Gp21 may be a structural protein. HHpred does not provide an alignment to PDB at a probability above 80%.
22	sheath	GeneID:2777386	
YP_006525.1	None	Genbank:YP_006525.1	HHpred identified the aa sequence as having a probability of 100.00 of being gp18 tail sheath protein of Enterobacteria phage T4 (PDB entry: 3J2M_Y). Gp22 is structural. PMID: 35458408
23	head morphogenesis	GeneID:2777387	
YP_006526.1	None	Genbank:YP_006526.1	No update necessary
parB	active partitioning of P1 plasmid prophage at cell division	GeneID:2777493	
YP_006527.1	None	Genbank:YP_006527.1	ParB has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification. In addition, parB has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. ParB may be a structural protein.
parA	active partitioning of P1 plasmid-prophage at cell division	GeneID:2777494	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006528.1	None	Genbank:YP_006528.1	ParA has been identified in moderate amounts by whole phage shotgun analysis using HPLC purified bands in addition to identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. HHpred identified the aa sequence as having a probability of 100.00 of being Plasmid partition protein A of Escherichia coli (PDB entry: 3EZ2_B). CDS has 100% identity to Plasmid partition protein A with 398 total aa spanning 1...398 aligned to aa 1...398 of Plasmid partition protein A. Various other bacterial plasmid partitioning proteins have also been identified with greater than 80 probability.
repA	plasmid replication	GeneID:2777439	
YP_006529.1	None	Genbank:YP_006529.1	HHpred identified the aa sequence as having a probability of 96.98 of being Replication initiation protein/DNA complex of Escherichia coli (PDB entry: 1REP_C). CDS has 96.98% identity to Replication initiation protein/DNA complex with 179 total aa spanning 86...270 aligned to aa 35...234 of Replication initiation protein/DNA complex.
upfA	unknown protein function	GeneID:2777397	
YP_006530.1	None	Genbank:YP_006530.1	Upfa has been significantly observed by whole phage shotgun analysis using HPLC purified bands, ammonium acetate triple purification, cesium chloride purification, and double cesium chloride purification. UpfA is structural protein. HHpred revealed no significant matches to PDB above 80% probability.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
mlp	membrane lipoprotein precursor	GeneID:2777398	
YP_006531.1	None	Genbank:YP_006531.1	HHpred identified the aa sequence as having a probability of 86.78 of being Outer membrane protein assembly factor of Escherichia coli K-12 (PDB entry: 6T1W_D). CDS has 32% identity to Outer membrane protein assembly factor with 19 total aa spanning 29...47 aligned to aa 5...23 of Outer membrane protein assembly factor.
ppfA	possible periplasmic function	GeneID:2777399	
YP_006532.1	None	Genbank:YP_006532.1	PpfA has been significantly identified by whole phage shotgun assembly by ammonium acetate triple purification, and cesium chloride purification. Additionally, ppfA has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions is structural though its role is unclear. PpfA is a structural protein though its role is unclear. HHpred revealed no significant matches to PDB above 80% probability.
upfB	unknown protein function	GeneID:2777400	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006533.1	None	Genbank:YP_006533.1	UpfB has been significantly identified by whole phage shotgun assembly by ammonium acetate triple purification and HPLC purified bands. Additionally, upfB has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virion. UpfB is a structural protein though its role is unclear. HHpred revealed no significant matches to PDB above 80% probability.
upfC	unknown protein function	GeneID:2777396	
YP_006534.1	None	Genbank:YP_006534.1	UpfC has been identified in low abundance by whole phage shotgun analysis using ammonium acetate triple purification. UpfC may be a structural protein.
uhr	gene upstream of hrdC	GeneID:2777401	
YP_006535.1	None	Genbank:YP_006535.1	HHpred revealed no significant matches to PDB above 80% probability.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
hrdC	homolog of rdgC	GeneID:2777402	
YP_006536.1	None	Genbank:YP_006536.1	HHpred identified the aa sequence as having a probability of 100.00 of being Recombination-associated protein rdgC of Pseudomonas aeruginosa (PDB entry: 2OWY_B). CDS has 40% identity to Recombination-associated protein rdgC with 296 total aa spanning 4...300 aligned to aa 2...302 of Recombination-associated protein rdgC. HrdC has been significantly observed by whole phage shotgun analysis using ammonium acetate triple purification, cesium chloride purification, and double cesium chloride purification. HrdC is structural.
dmt	DNA methylation	GeneID:2777403	
YP_006537.1	None	Genbank:YP_006537.1	Dmt has been identified in a low amount by whole phage shotgun analysis using double cesium chloride purification in addition to identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Dmt may be structural.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
trnA	tRNA-Asn	GeneID:2777391	
trnT	tRNA-Thr	GeneID:2777392	
plp	putative lipoprotein, by homology	GeneID:2777393	
YP_006538.1	None	Genbank:YP_006538.1	HHpred identified the aa sequence as having a probability of 81.69 of being EGF family domain-containing protein of Toxoplasma gondii (PDB entry: 4Z80_C). CDS has 30% identity to EGF family domain-containing protein with 19 total aa spanning 41...60 aligned to aa 6...25 of EGF family domain-containing protein.
upl	gene upstream of plp	GeneID:2777394	
YP_006539.1	None	Genbank:YP_006539.1	HHpred does not provide an alignment to PDB at a probability above 80%.
tciA	tellurite or colicin resistnce or inhibition of cell division, by homology	GeneID:2777433	
YP_006540.1	None	Genbank:YP_006540.1	TciA has been significantly observed by whole phage shotgun analysis using ammonium acetate triple purification and HPLC purified bands. Additionally, tciA has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virion. HHpred identified the aa sequence as having a probability of 97.7 of being TerB of Klebsiella pneumoniae (PDB entry: 2JXU_A). CDS has 15% identity to TerB with 128 total aa spanning 12…140 aligned to aa 18…149 of TerB. TciA is a structural protein.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
tciB	tellurite or colicin resistance or inhibition of cell division accessory function, by homology	GeneID:2777389	
YP_006541.1	None	Genbank:YP_006541.1	HHpred does not provide an alignment to PDB at a probability above 80%.
tciC	tellurite or colicin resistance or inhibition of cell division accessory protein	GeneID:2777390	
YP_006542.1	None	Genbank:YP_006542.1	HHpred does not provide an alignment to PDB at a probability above 80%.
trnI	tRNA-Ile	GeneID:2777467	
ban	dnaB analog and homolog	GeneID:2777486	
YP_006543.1	None	Genbank:YP_006543.1	Ban has been identified in a low amount by whole phage shotgun analysis using ammonium acetate triple purification in addition to identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. HHpred identified the aa sequence as having a probability of 100.00 of being Replicative DNA helicase (E.C.3.6.4.12) of Escherichia coli O111:NM (PDB entry: 6BBM_D). CDS has 79% identity to Replicative DNA helicase (E.C.3.6.4.12) with 443 total aa spanning 6...448 aligned to aa 24...466 of Replicative DNA helicase (E.C.3.6.4.12). Ban may be structural.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
dbn	gene downstream of ban	GeneID:2777422	
YP_006544.1	None	Genbank:YP_006544.1	Dbn has been significantly identified by whole phage shotgun assembly by ammonium acetate triple purification. HHpred identified the aa sequence as having a probability of 96.41 of being Glycosyltransferase C of Streptococcus agalactiae COH1 (PDB entry: 4W6Q_B). CDS has 8% identity to Glycosyltransferase C with 96 total aa spanning 227...322 aligned to aa 214...330 of Glycosyltransferase. Dbn is structural.
5	baseplate	GeneID:2777423	
YP_006545.1	None	Genbank:YP_006545.1	Gp5 has been significantly identified by whole phage shotgun assembly by ammonium acetate triple purification and cesium chloride purification. HHpred identified the aa sequence as having a probability of 99.67 of being gp5 tail needle of Enterobacteria phage T4 (PDB entry: 1WTH_A). Gp5 is structural. PMID: 35458408
6	tail length	GeneID:2777424	
YP_006546.1	None	Genbank:YP_006546.1	HHpred identified the aa sequence as having a probability of 99.06 of being gp27 baseplate structure protein of Enterobacteria phage T4 (PDB entry: 1WTH_D). Gp6 is structural. PMID: 35458408
24	baseplate or tail stability	GeneID:2777425	
YP_006547.1	None	Genbank:YP_006547.1	HHpred identified the aa sequence as having a probability of 94.58 of being gp15 tail connector protein of Enterobacteria phage T4 (PDB entry: 4HUD_C). Gp24 is structural. PMID: 35458408

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
7	tail stability	GeneID:2777426	
YP_006548.1	None	Genbank:YP_006548.1	Gp7 has been identified in significant abundance by whole phage shotgun analysis using ammonium acetate triple purification and double cesium chloride purified virions. In addition, hxr has been identification in significant abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Gp7 is structural though its role is unclear. HHpred does not provide an alignment to PDB at a probability above 80%. PMID: 35458408
25	tail stability	GeneID:2777427	
YP_006549.1	None	Genbank:YP_006549.1	Gp25 has been identified in significant abundance by whole phage shotgun analysis using ammonium acetate triple purification and cesium chloride purified virions. In addition, hxr has been identification in significant abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Gp25 is structural though its role is unclear. HHpred does not provide an alignment to PDB at a probability above 80%. PMID: 35458408
26	baseplate	GeneID:2777428	
YP_006550.1	None	Genbank:YP_006550.1	HHpred identified the aa sequence as having a probability of 98.9 of being gp53 wedge component of Enterobacteria phage T4 (PDB entry: 5HX2_F). Gp26 is structural. PMID: 35458408
pmgL	putative morphogenetic function	GeneID:2777430	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006551.1	None	Genbank:YP_006551.1	HHpred identified the aa sequence as having a probability of 94.9 of being DNA-directed RNA polymerase subunit alpha of Escherichia coli (strain K12) (PDB entry: 5W1S_N). CDS has 21% identity to DNA-directed RNA polymerase subunit alpha with 52 total aa spanning 12...65 aligned to aa 9...68 of DNA-directed RNA polymerase subunit alpha.
pmgM	putative morphogenetic function	GeneID:2777447	
YP_006552.1	None	Genbank:YP_006552.1	HHpred identified the aa sequence as having a probability of 99.49 of being VRR-NUC of Psychrobacter sp. (PDB entry: 4QBL_A). CDS has 35% identity to VRR-NUC with 124 total aa spanning 59…182 aligned to aa 5…141 of VRR-NUC.
pmgN	putative morphogenetic function	GeneID:2777448	
YP_006553.1	None	Genbank:YP_006553.1	HHpred does not provide an alignment to PDB at a probability above 80%.
pmgO	putative morphogenetic function	GeneID:2777449	
YP_006554.1	None	Genbank:YP_006554.1	PmgO has been identified in a low amount by whole phage shotgun analysis using ammonium acetate triple purification. HHpred identified the aa sequence as having a probability of 98.25 of being alpha-2,3/8-sialyltransferase (E.C.2.4.99.-) of Campylobacter jejuni (PDB entry: 1RO7_C). CDS has 15% identity to alpha-2,3/8-sialyltransferase (E.C.2.4.99.-) with 176 total aa spanning 20…198 aligned to aa 2…230 of alpha-2,3/8-sialyltransferase (E.C.2.4.99.-). PmgO may be structural.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
pmgP	putative morphogenetic function	GeneID:2777420	
YP_006555.1	None	Genbank:YP_006555.1	PmgP has been identified in significant abundance by whole phage shotgun analysis using HPLC purification, ammonium acetate triple purification, cesium chloride purification, and double cesium chloride purified virions. HHpred identified the aa sequence as having a probability of 97.90 of being alpha-2,3/8-sialyltransferase (E.C.2.4.99.-) of Campylobacter jejuni (PDB entry: 2P2V_A). CDS has 17% identity to alpha-2,3/8-sialyltransferase (E.C.2.4.99.-) with 143 total aa spanning 9…163 aligned to aa 18…180 of alpha-2,3/8-sialyltransferase (E.C.2.4.99.-). PmgP is a structural protien.
ppp	P1 protein phosphatase	GeneID:2777421	
YP_006556.1	None	Genbank:YP_006556.1	Ppp has been identified in significant abundance by whole phage shotgun analysis using HPLC. Ppp is a structural protein.
pmgQ	putative morphogenetic function	GeneID:2777379	
YP_006557.1	None	Genbank:YP_006557.1	PmgQ has been identified in significant abundance by whole phage shotgun analysis using ammonium acetate triple purification. HHpred aligns pmgQ to proteins of the kingdom animalia. PmgQ may be structural.
pmgR	putative morphogenetic function	GeneID:2777455	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006558.1	None	Genbank:YP_006558.1	PmgR has been identified in a low amount by whole phage shotgun analysis using ammonium acetate triple purification. HHpred does not provide an alignment to PDB at a probability above 80%.
pmgS	putative morphogenetic function	GeneID:2777456	
YP_006559.1	None	Genbank:YP_006559.1	PmgS has been identified in significant abundance by whole phage shotgun analysis using HPLC purified bands, ammonium acetate triple purification, cesium chloride purified virions, and double cesium chloride purified virions. In addition, pmgS has been identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. PmgS is a structural protein though its structural role is unclear. HHpred does not provide an alignment to PDB at a probability above 80%.
pap	P1 acid phosphatase	GeneID:2777457	
YP_006560.1	None	Genbank:YP_006560.1	HHpred identified the aa sequence as having a probability of 97.15 of being Polynucleotide kinase (E.C.2.7.1.78,3.1.3.34) of Enterobacteria phage T4 (PDB entry: 5UJ0_A). CDS has 25% identity to Polynucleotide kinase (E.C.2.7.1.78,3.1.3.34) with 129 total aa spanning 12…157 aligned to aa 2…140 of Polynucleotide kinase (E.C.2.7.1.78,3.1.3.34).
pmgT	putative morphogenetic function	GeneID:2777438	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006561.1	None	Genbank:YP_006561.1	PmgT has been identified in a low amount by whole phage shotgun analysis using HPLC purified bands in addition to identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. HHpred identified the aa sequence as having a probability of 96.28 of being ASCH domain of Enterococcus faecalis (PDB entry: 3S9X_A). CDS has 24% identity to ASCH domain with 67 total aa spanning 159…236 aligned to aa 90…158 of ASCH domain.
pmgU	putative morphogenetic function	GeneID:2777445	
YP_006562.1	None	Genbank:YP_006562.1	HHpred does not provide an alignment to PDB at a probability above 80%.
pmgV	putative moprhogenetic function	GeneID:2777446	
YP_006563.1	None	Genbank:YP_006563.1	PmgT has been identified in a low amount by whole phage shotgun analysis using HPLC purified bands. HHpred does not provide an alignment to PDB at a probability above 80%.
upfM	unknown protein function	GeneID:2777440	
YP_006564.1	None	Genbank:YP_006564.1	HHpred does not provide an alignment to PDB at a probability above 80%.
upfN	unknown protein function	GeneID:2777441	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006565.1	None	Genbank:YP_006565.1	HHpred does not provide an alignment to PDB at a probability above 80%.
upfO	unknown protein function	GeneID:2777442	
YP_006566.1	None	Genbank:YP_006566.1	HHpred does not provide an alignment to PDB at a probability above 80%.
hot	homolog of gene for theta subunit of *E. coli* DNA polymerase III	GeneID:2777458	
YP_006567.1	None	Genbank:YP_006567.1	Hot has been identified in a low amount by whole phage shotgun analysis using double cesium chloride purified virions in addition to identification in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions.
lxr	LexA regulated function	GeneID:2777459	
YP_006568.1	None	Genbank:YP_006568.1	HHpred does not provide an alignment to PDB at a probability above 80%.
humD	homolog of umuD	GeneID:2777472	
YP_006569.1	None	Genbank:YP_006569.1	No update necessary
phd	prevention of host death by Doc toxicity	GeneID:2777473	

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
YP_006570.1	None	Genbank:YP_006570.1	Phd has been identified in a low amount by whole phage shotgun analysis using ammonium acetate triple purification. Phd may be a structural protein.
doc	death on curing	GeneID:2777474	
YP_006571.1	None	Genbank:YP_006571.1	No update necessary
pdcA	post-doc	GeneID:2777475	
YP_006572.1	None	Genbank:YP_006572.1	HHpred does not provide an alignment to PDB at a probability above 80%.
pdcB	post-doc	GeneID:2777483	
YP_006573.1	None	Genbank:YP_006573.1	HHpred identified the aa sequence as having a probability of 96.28 of being prophage-derived uncharacterized protein ybcO of Escherichia coli K-12 (PDB entry: 3G27_A). CDS has 24% identity to prophage-derived uncharacterized protein ybcO with 85 total aa spanning 257…341 aligned to aa 2…96 of prophage-derived uncharacterized protein ybcO.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
lpa	late promoter activation	GeneID:2777484	
YP_006574.1	None	Genbank:YP_006574.1	Ipa has been identified in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. Ipa may be structural. HHpred does not provide an alignment to PDB at a probability above 80%.
pacA	phage DNA packaging	GeneID:2777434	
YP_006575.1	None	Genbank:YP_006575.1	Hot has been identified in a low amount by whole phage shotgun analysis using ammonium acetate triple purification and cesium chloride double purification of virions. PacA may be structural though there is no precedent in model phages where the terminase is structural.
pacB	phage DNA packaging	GeneID:2777432	
YP_006576.1	None	Genbank:YP_006576.1	No update necessary
c1.100	phage repressor ts mutant c1.100	GeneID:2777431	
YP_006577.1	None	Genbank:YP_006577.1	C1.100 has been identified in significant abundance by whole phage shotgun analysis using HPLC purified bands, ammonium acetate triple purification, cesium chloride purification, and cesium chloride double purification of virions. C1.100 is structural.

Table 8 Continued

Name	Note first clause	DBxrefs	CDS Comment Update
coi	C one inactivation	GeneID:2777460	
YP_006578.1	None	Genbank:YP_006578.1	HHpred does not provide an alignment to PDB at a probability above 80%.
imcB	immunity C function	GeneID:2777461	
YP_006579.1	None	Genbank:YP_006579.1	ImcB was identified in low abundance by LC-MS/MS of cesium chloride isopycnic ultra centrifugation purified P1 virions. HHpred does not provide an alignment to PDB at a probability above 80%.
imcA	immunity C function	GeneID:2777462	
YP_006580.1	None	Genbank:YP_006580.1	HHpred does not provide an alignment to PDB at a probability above 80%.